

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application: 2 0 0 4 年 2 月 6 日

出 願 番 号
Application Number: 特 願 2 0 0 4 - 0 3 1 1 5 0
[ST. 10/C]: [J P 2 0 0 4 - 0 3 1 1 5 0]

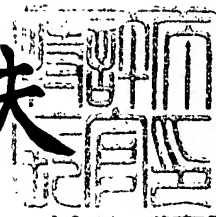
出 願 人
Applicant(s): 株式会社日立製作所

CERTIFIED COPY OF
PRIORITY DOCUMENT

2 0 0 4 年 4 月 2 1 日

特許庁長官
Commissioner,
Japan Patent Office

今 井 康 夫



出証番号 出証特 2 0 0 4 - 3 0 3 4 4 9 2

【書類名】 特許願
【整理番号】 340301787
【あて先】 特許庁長官殿
【国際特許分類】 G06F 03/06
【発明者】
 【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A I
 D システム事業部内
 【氏名】 渡辺 治明
【発明者】
 【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A I
 D システム事業部内
 【氏名】 占部 喜一郎
【発明者】
 【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A I
 D システム事業部内
 【氏名】 山神 憲司
【特許出願人】
 【識別番号】 000005108
 【氏名又は名称】 株式会社日立製作所
【代理人】
 【識別番号】 100095371
 【弁理士】
 【氏名又は名称】 上村 輝之
【選任した代理人】
 【識別番号】 100089277
 【弁理士】
 【氏名又は名称】 宮川 長夫
【選任した代理人】
 【識別番号】 100104891
 【弁理士】
 【氏名又は名称】 中村 猛
【手数料の表示】
 【予納台帳番号】 043557
 【納付金額】 21,000円
【提出物件の目録】
 【物件名】 特許請求の範囲 1
 【物件名】 明細書 1
 【物件名】 図面 1
 【物件名】 要約書 1
 【包括委任状番号】 0110323

【書類名】 特許請求の範囲**【請求項 1】**

ホスト端末に接続された記憶制御システムが有する記憶制御サブシステムにおいて、データを論理的に記憶する論理記憶デバイスと、前記論理記憶デバイスを有し、前記論理的に記憶されたデータを物理的に記憶する物理記憶デバイスと、

仮想記憶容量値が設定されることによって実装状態となる仮想記憶ユニットと、

前記設定された仮想記憶容量値を記憶するメモリと、

前記仮想記憶ユニットを認識している前記ホスト端末からリード要求又はライト要求を受信した場合に、前記仮想記憶ユニットにおける仮想記憶領域と、前記論理記憶デバイスにおける論理記憶領域とを対応付けて、前記仮想記憶領域を介して前記論理記憶領域と前記ホスト端末との間でデータをやり取りする記憶制御部とを備え、

前記記憶制御部が、前記メモリに記憶された仮想記憶容量値を前記ホスト端末に通知し、前記ホスト端末において前記仮想記憶容量値が記憶された後、前記仮想記憶ユニットが前記ホスト端末に接続されている間は、前記通知した仮想記憶容量値が変更されないようにする、

記憶制御サブシステム。

【請求項 2】

前記記憶制御サブシステムの保守のための処理を行う保守用端末が前記記憶制御部に接続されている場合において、

前記記憶制御部は、前記保守用端末又は前記保守用端末に接続された外部端末から、新たな前記仮想記憶ユニットを準備することのユニット準備要求を受け、前記ユニット準備要求に応答して、少なくとも前記仮想記憶容量値の記入欄を持ったグラフィカルユーザインターフェースを前記保守用端末又は前記外部端末に提供し、前記記入欄に入力された仮想記憶容量値を、前記設定された仮想記憶容量値として前記メモリに記憶させる、請求項 1 記載の記憶制御サブシステム。

【請求項 3】

前記記憶制御サブシステムが、2つの記憶ユニットから成るユニットペアを形成し、一方の記憶ユニットをプライマリ記憶ユニット、他方の記憶ユニットをセカンダリ記憶ユニットとして、プライマリ記憶ユニット内のデータをセカンダリ記憶ユニットにコピーするスナップショットを行うようになっている場合、

前記物理記憶デバイスには複数の前記論理記憶デバイスが設けられており、

前記複数の論理記憶デバイスには、前記仮想記憶領域に対応付けられ得る論理記憶領域を持った2以上の第1論理記憶デバイスと、前記仮想記憶領域に対応付けられることのない論理記憶領域を持った1以上の第2論理記憶デバイスとが含まれており、

前記1以上の第2論理記憶デバイスが、前記ホスト端末に接続される1つのリアル記憶ユニットを構成し、

前記記憶制御部は、前記リアル記憶ユニットを前記プライマリ記憶ユニットとし、前記仮想記憶ユニットをセカンダリ記憶ユニットとしたユニットペアを形成して前記スナップショットを行う、

請求項 1 記載の記憶制御サブシステム。

【請求項 4】

前記記憶制御部は、前記仮想記憶容量値を前記ホスト端末に通知していない場合に、前記仮想記憶ユニットと前記リアル記憶ユニットの前記ユニットペアを形成するならば、前記リアル記憶ユニットの記憶容量値と同じ値を前記仮想記憶ユニットの記憶容量値として前記ホスト端末に通知する、

請求項 3 記載の記憶制御サブシステム。

【請求項 5】

前記記憶制御部は、前記仮想記憶ユニットの相手となる前記リアル記憶ユニットが見つ

からない場合に、前記ホスト端末から前記仮想記憶ユニットに対してリード要求又はライト要求を受けたならば、前記仮想記憶ユニットが未実装状態であると前記ホスト端末に通知し、その後、前記相手となるリアル記憶ユニットが見つかったならば、前記リアル記憶ユニットの記憶容量値と同じ値を前記仮想記憶ユニットの記憶容量値として前記ホスト端末に通知する、

請求項 4 記載の記憶制御サブシステム。

【請求項 6】

前記記憶制御部は、前記ホスト端末から所定コマンドを受信した場合に、前記メモリに記憶されている仮想記憶容量値を前記ホスト端末に通知する、

請求項 1 記載の記憶制御サブシステム。

【請求項 7】

前記記憶制御サブシステムの保守のための処理を行う保守用端末が前記記憶制御部に接続されており、且つ、前記ホスト端末が行う所定の処理で、前記ホスト端末が、前記記憶した仮想記憶容量値を消去するようになっている場合において、

前記記憶制御部は、前記仮想記憶容量値を前記ホスト端末に通知した後に、前記保守用端末又は前記保守用端末に接続された外部端末から、前記通知した仮想記憶容量値の更新要求を受けたならば、前記ホスト端末と前記仮想記憶ユニットとが接続されていない間に、前記保守用端末又は前記外部端末から新たな仮想記憶容量値を受けて前記メモリに記憶させ、且つ、前記ホスト端末に前記所定の処理を行わせることで、前記ホストに記憶されている旧い前記仮想記憶容量値が消去されるようにした後、前記メモリに記憶させた新たな仮想記憶容量値を前記ホスト端末に通知する、

請求項 1 記載の記憶制御サブシステム。

【書類名】明細書

【発明の名称】仮想記憶ユニットを備えた記憶制御サブシステム

【技術分野】

【0001】

本発明は、仮想記憶ユニットを備えた記憶制御サブシステムに関する。

【背景技術】

【0002】

例えば、大容量のデータを取り扱う基幹業務用の記憶システムでは、ホストコンピュータ（以下、「ホスト端末」と言う）とは、別体に構成された記憶制御サブシステムを用いてデータが管理されている。この記憶制御サブシステムは、例えば、多数のディスク型記憶装置をアレイ状に配置して構成されている R A I D（Redundant Array of Independent Inexpensive Disks）である。

【0003】

例えば、特開平 9-288547 号公報には、直列に並べられた複数の実デバイスを 1 つの仮想デバイスとして提供することが開示されている（段落 23）。

【0004】

また、例えば、特開平 11-345158 号公報には、正副ボリュームへの更新を各ボリュームの更新情報にそれぞれ記録する正副 1 対のボリュームに対する差分ダンプ取得方法であって、ミラー状態移行時に前記副ボリュームの更新情報を前記正ボリュームの更新情報に一致させ、ダンプ開始時に前記ミラー状態からミラー一時解除状態へ移行して前記正ボリュームの更新情報を初期化するとともに前記副ボリュームから更新部分のダンプを得ることが開示されている。

【0005】

【特許文献 1】特開平 9-288547 号公報。

【特許文献 2】特開平 11-345158 号公報。

【発明の開示】

【発明が解決しようとする課題】

【0006】

ところで、例えば、記憶制御サブシステムは、1 又は複数のディスク型記憶装置上に用意される複数の論理的な記憶デバイス（以下、Logical Device を略して「L D E V」と言う）のうち、1 以上の L D E V から構成される 1 つの記憶ユニットを L U（Logical Unit）としてホスト端末に提供するものがある。このような記憶制御システムでは、例えば、1 つの L U をプライマリ L U、別の 1 つの L U をセカンダリ L U とした L U ペアが形成されて、プライマリ L U 内の更新前データをセカンダリ L U にコピーすること（すなわち、いわゆるスナップショットを作成すること）が行われる場合がある。

【0007】

一般に、L U ペアを構成する双方の L U の記憶容量は同じである。このため、プライマリ L U の記憶容量が大きいと、セカンダリ L U としても記憶容量の大きい L U が必要となる。しかし、上述したスナップショットでは、コピーされる更新前データのデータサイズが小さい場合がある。そのため、セカンダリ L U の記憶容量は無駄に大きいことになってしまう場合がある。

【0008】

また、例えば、記憶制御サブシステムは、ホスト端末からリードキャパシティコマンドを受けた場合、その記憶制御サブシステムが有する L U の記憶容量をホスト端末に通知することがある。ここで、記憶制御サブシステムが、同一の L U について、過去に通知した記憶容量と別の記憶容量を通知すると、ホスト端末に混乱を生じさせてしまう可能性がある。

【0009】

従って、本発明は、例えば、少なくとも次のいずれかの目的を達成するものである。

(1) スナップショット作成のために、記憶容量の大きい L U ペアを形成しても、無駄な空

き記憶領域を少なくする記憶制御サブシステムを提供する。

(2) ホスト端末に混乱を生じさせてしまう可能性を低減する記憶制御サブシステムを提供する。

【0010】

本発明の他の目的は、後述の説明から明らかになるであろう。

【課題を解決するための手段】

【0011】

本発明に従う記憶制御サブシステムは、ホスト端末に接続された記憶制御システムが有する記憶制御サブシステムであって、データを論理的に記憶する論理記憶デバイスと、前記論理記憶デバイスを有し、前記論理的に記憶されたデータを物理的に記憶する物理記憶デバイスと、仮想記憶容量値が設定されることによって実装状態となる仮想記憶ユニットと、前記設定された仮想記憶容量値を記憶するメモリと、記憶制御部とを備える。記憶制御部は、前記仮想記憶ユニットを認識している前記ホスト端末からリード要求又はライト要求を受信した場合に、前記仮想記憶ユニットにおける仮想記憶領域と、前記論理記憶デバイスにおける論理記憶領域とを対応付けて（例えば動的に対応付けて）、前記仮想記憶領域を介して前記論理記憶領域と前記ホスト端末との間でデータをやり取りする。前記記憶制御部は、前記メモリに記憶された仮想記憶容量値を前記ホスト端末に通知し、前記ホスト端末において前記仮想記憶容量値が記憶された後、前記仮想記憶ユニットが前記ホスト端末に接続されている間は、前記通知した仮想記憶容量値が変更されないようにする。

【0012】

本発明に従う記憶制御方法は、例えば、データを論理的に記憶する論理記憶デバイスと、前記論理記憶デバイスを有し、前記論理的に記憶されたデータを物理的に記憶する物理記憶デバイスと、仮想記憶容量値が設定されることによって実装状態となる仮想記憶ユニットと、前記設定された仮想記憶容量値を記憶するメモリとを備える記憶制御サブシステムにおいて行われる方法である。前記記憶制御方法は、前記仮想記憶ユニットを認識している前記ホスト端末からリード要求又はライト要求を受信した場合に、前記仮想記憶ユニットにおける仮想記憶領域と、前記論理記憶デバイスにおける論理記憶領域とを対応付けるステップと、前記仮想記憶領域を介して前記論理記憶領域と前記ホスト端末との間でデータをやり取りするステップと、前記メモリに記憶された仮想記憶容量値を前記ホスト端末に通知するステップと、前記通知するステップを実行し、前記ホスト端末において前記仮想記憶容量値が記憶された後、前記仮想記憶ユニットが前記ホスト端末に接続されている間は、前記通知した仮想記憶容量値が変更されないようにするステップとを有する。

【発明の効果】

【0013】

本発明によれば、ホスト端末に混乱を生じさせてしまう可能性を低減する記憶制御サブシステムが提供される。

【発明を実施するための最良の形態】

【0014】

以下、図面を参照して本発明の一実施形態を説明する。

【0015】

図1は、本発明の一実施形態に係る記憶制御システムの外觀の概略を示す。

【0016】

記憶制御システム600は、例えば、基本筐体10と複数の増設筐体12とから構成することができる（基本筐体11のみで構成されても良い）。

【0017】

基本筐体10は、記憶制御システム600の最小構成単位である。この基本筐体10には、例えば、複数のディスク型記憶装置（例えばハードディスクドライブ（HDD））300と、複数の制御パッケージ（例えば後述するチャネル制御部又はディスク制御部）105と、複数の電源ユニット400と、複数のバッテリーユニット500とがそれぞれ着脱可能に設けられている。また、基本筐体10には、複数の冷却ファン13が設けられてい

る。

【0018】

各増設筐体12は、記憶制御システム600のオプションであり、例えば、1つの基本筐体10に最大4個の増設筐体12を接続することができる。各増設筐体12には、複数の冷却ファン13が設けられている。また、各増設筐体12には、複数のディスク型記憶装置300と、複数の電源ユニット400と、複数のバッテリーユニット500とがそれぞれ着脱可能に設けられており、それらの各々は、例えば、基本筐体10に設けられた制御パッケージ105が有する制御機能により制御される。

【0019】

図2は、本発明の一実施形態に係る記憶システムの全体構成例を示す。

【0020】

この記憶システム1の基本的な構成要素は、1又は複数のホスト端末200A～200Dと、記憶制御システム600である。

【0021】

ホスト端末（上位装置）200A～200Dの各々は、例えば、CPU（Central Processing Unit）、不揮発性及び／又は揮発性のメモリ（例えばROM又はRAM）、及びハードディスク等をハードウェア資源として備えたコンピュータシステム（例えば、パーソナルコンピュータ又はワークステーション）である。各ホスト端末200A～200DのCPUが、メモリに格納された各種コンピュータプログラムを読込んで実行することにより、コンピュータプログラムとハードウェア資源（例えばメモリ）とが協働した処理が行われて、種々の機能が実現される。ホスト端末200A～200Dの各々は、種々の方法で記憶制御システム600に接続することができる。

【0022】

例えば、ホスト端末200A～200Bは、第一通信ネットワーク（例えば、LAN、インターネット又は専用回線、以下、LANであるとする）820を介して記憶制御システム600に接続されている。LAN820を介して行われるホスト端末200A及び200Bと記憶制御システム600との間の通信は、例えばTCP/IPプロトコルに従って行われる。ホスト端末200A及び200Bから記憶制御システム600に対して、ファイル名指定によるデータアクセス要求（ファイル単位でのデータ入出力要求、以下、「ファイルアクセス要求」と言う）が送信される。

【0023】

また、例えば、ホスト端末200B及び200Cは、第二通信ネットワーク（例えば、SAN（Storage Area Network）、以下、SANであるとする）821を介して記憶制御システム600に接続されている。SAN821を介して行われるホスト端末200B及び200Cと記憶制御システム600との間の通信は、例えばファイバチャネルプロトコルに従って行われる。ホスト端末200B及び200Cから記憶制御システム600に対して、例えば、ブロック単位のデータアクセス要求（以下、「ブロックアクセス要求」と言う）が送信される（なお、ブロック単位とは、後述のディスク型記憶装置300上の記憶領域におけるデータの管理単位であるブロックを単位としたものである）。

【0024】

また、例えば、ホスト端末200Dは、LAN820やSAN821等のネットワークを介さずに記憶制御システム600に接続されている。ホスト端末200Dは、例えば、メインフレームコンピュータとすることができる。ホスト端末200Dと記憶制御システム600との間の通信は、例えば、FICON（Fibre Connection：登録商標）、ESCON（Enterprise System Connection：登録商標）、ACONARC（Advanced Connection Architecture：登録商標）、FIBARC（Fibre Connection Architecture：登録商標）等の通信プロトコルに従って行われる。ホスト端末200Dから記憶制御システム600に対して、例えば、これらの通信プロトコルのうちのいずれかに従ってブロックアクセス要求が送信される。

【0025】

LAN 820 及び SAN 821 の少なくとも一方には、例えば、バックアップ記憶制御システム 910 が接続されている。バックアップ記憶制御システム 910 は、例えば、1 又は複数のディスク系デバイス（例えば、MO、CD-R、又は DVD-RAM）から選択されたディスク系デバイスにデータを格納する記憶制御システムであっても良いし、後に詳述する記憶制御システム 600 のような記憶制御システムであっても良いし、1 又は複数のテープ系デバイス（例えば、DAT テープ、カセットテープ、オープンテープ又はカートリッジテープ）から選択されたテープ系デバイスにデータを格納する記憶制御システムであっても良い。バックアップ記憶制御システム 910 は、例えば、LAN 820 或いは SAN 821（又は、更に、バックアップ制御システム 910 に接続されているホスト端末）を介して、記憶制御システム 600 に記憶されているデータを受信し、自分が備えている記憶装置（例えばテープ系デバイス）にそのデータを格納する。

【0026】

また、LAN 820 及び SAN 821 のうちの少なくとも LAN 820 には、例えば、管理サーバ 819 が接続されていても良い。管理サーバ 819 は、例えば、ホスト端末と別のホスト端末との間の通信や、ホスト端末と記憶制御システム 600 との間の通信の中継を行っても良い。

【0027】

また、ホスト端末 200A~200D は、第三通信ネットワーク（例えば LAN）を介して相互に接続されていても良い。

【0028】

記憶制御システム 600 は、例えば、RAID（Redundant Array of Independent Inexpensive Disks）システムである。記憶制御システム 600 は、ホスト端末 200A~200D から受信したコマンドに従う制御を行う。記憶制御システム 600 は、記憶制御サブシステム（ディスクアレイ装置）102 と、保守用端末（以下、Service Processor を略して「SVP」と記載）160 とを備える。記憶制御サブシステム 102 は、記憶制御装置 100 と、記憶装置ユニット 101 とを備える。記憶制御装置 100 は、1 又は複数のチャンネル制御部 110A~110D と、1 又は複数のキャッシュメモリ（以下、Cache Memory を略して「CM」と記載）130 と、1 又は複数の共有メモリ（以下、Shared Memory を略して「SM」と記載）120 と、1 又は複数のディスク制御部 140A~140D と、接続部 150 とを備える。記憶装置ユニット 101 は、1 以上の物理ディスク群 5 を備える。1 以上の物理ディスク群 5 の各々は、アレイ状に配列された複数のディスク型記憶装置 300 を有する。

【0029】

チャンネル制御部 110A~110D の各々は、ハードウェア回路、ソフトウェア、又はそれらの組み合わせで構成することができる。各チャンネル制御部 110A~110D は、記憶制御装置 100（例えば基本筐体 10）に対して着脱可能であり、チャンネルアダプタと呼ばれることがある。各チャンネル制御部 110A~110D は、例えば、プロセッサやメモリ等が実装されたプリント基板と、メモリに格納された制御プログラムとをそれぞれ備えており、これらのハードウェアとソフトウェアとの協働作業によって、所定の機能を実現する。チャンネル制御部 110A~110D の各々は、多重化（例えば二重化）されており、1 つのチャンネル制御部が破損しても他のチャンネル制御部で動作するようになっている。チャンネル制御部 110A~110D は、SM 120 内の制御情報等（例えば後述の L U-L D E V 管理テーブル）を参照しつつ、ホスト端末から受信したコマンドに従う処理を実行する。チャンネル制御部 110A~110D のうち、チャンネル制御部 110C を例に採り、ディスク制御部 140A~140D の動作も含めて先に説明すると、例えば、チャンネル制御部 110C は、ホスト端末 200A 又は 200B から、リード要求を含んだ I/O 要求（入/出力要求、ここではブロックアクセス要求）を受信すると、読出しコマンドを SM 120 に記憶させると共に、CM 130 にキャッシュ領域を確保する。ディスク制御部 140A~140D は、SM 120 を随時参照しており、未処理の読出しコマンドを発見すると、ディスク型記憶装置 300 からデータ（典型的には、ホスト端末 200A~

200Dとディスク型記憶装置300との間でやり取りされるユーザデータ)を読み出して、CM130に確保された上記キャッシュ領域に記憶させる。チャンネル制御部110Cは、CM130に移されたデータをキャッシュ領域から読み出し、そのデータをリード要求発行元のホスト端末200A又は200Bに送信する。

【0030】

また、例えば、チャンネル制御部110Cは、ホスト端末200B又は200Cから、ライト要求を含んだI/O要求を受信すると、書込みコマンドをSM120に記憶させると共に、CM130にキャッシュ領域を確保し、受信したI/O要求に含まれているデータを、上記確保したキャッシュ領域に記憶させる。その後、チャンネル制御部110Cは、ライト要求発行元のホスト端末200B又は200Cに対して書込み完了を報告する。そして、ディスク制御部140A~140Dは、SM120を随時参照しており、未処理の書込みコマンドを発見すると、その書込みコマンドに従って、CM130に確保された上記キャッシュ領域からデータを読み出し、そのデータを所定のディスク型記憶装置300に記憶させる。

【0031】

上記の処理は、他のチャンネル制御部110A~110B及び110Dも行うことができる。なお、チャンネル制御部110A~110Bは、ファイルアクセス要求をブロックアクセス要求に変換した後に(例えば、自分が持っているファイルシステムに従い、ファイル要求に含まれているファイル名を論理ブロックアドレスに変換した後に)、上記の処理を行う。

【0032】

ディスク制御部140A~140Dは、ハードウェア回路、ソフトウェア、又はそれらの組み合わせで構成することができる。各チャンネル制御部140A~140Dは、記憶制御装置100(例えば基本筐体10又は増設筐体12)に対して着脱可能であり、ディスクアダプタと呼ばれることがある。ディスク制御部140A~140Dは、例えば、プロセッサやメモリ等が実装されたプリント基板と、メモリに格納された制御プログラムとをそれぞれ備えており、これらのハードウェアとソフトウェアとの協働作業によって、所定の機能を実現する。ディスク制御部140A~140Dの各々は、多重化(例えば二重化)されており、1つのディスク制御部が破損しても他のディスク制御部で動作するようになっている。ディスク制御部140A~140Dは、SM120内の制御情報等(例えば後述のLU-LDEV管理テーブル)を参照しつつ、各物理ディスク群5に含まれる各ディスク型記憶装置300との間のデータ通信を制御するものである。各ディスク制御部140A~140Dと各ディスク型記憶装置300とは、例えば、SAN等の通信ネットワークを介して接続されており、ファイバチャネルプロトコルに従ってブロック単位のデータ転送を行う。また、ディスク制御部140A~140Dは、ディスク型記憶装置300の状態を随時監視しており、この監視結果は内部の通信ネットワーク(例えばLAN)151を介してSVP160に送信される。

【0033】

1又は複数のCM130は、例えば、揮発性又は不揮発性のメモリである。CM130には、キャッシュ領域が確保され、そこに、チャンネル制御部110A~110Dとディスク制御部140A~140Dとの間で送受されるデータが記憶される。そのデータは、複数のCM130により多重管理されても良い。

【0034】

1又は複数のSM120は、例えば不揮発性のメモリから構成され、制御情報等を記憶する(例えば、制御情報等は、複数のSM120により多重管理されても良い)。制御情報等には、例えば、チャンネル制御部110A~110Dとディスク制御部140A~140Dとの間でやり取りされる種々のコマンドや、キャッシュ管理テーブルや、ディスク管理テーブルや、LU-LDEV管理テーブルがある。キャッシュ管理テーブルは、例えば、キャッシュ領域と、後述するLDEVの論理アドレスとの対応関係が書かれたテーブルである。ディスク管理テーブルは、各ディスク型記憶装置300を管理するためのテーブ

ルであり、例えば、各ディスク型記憶装置300毎に、ディスクID、ペンダ、記憶容量、RAIDレベル、使用状況（例えば使用中か未使用か）等を有する。LU-LEDEV管理テーブルは、後述するLEDEVを管理するためのテーブルであり、例えば、各LEDEV毎に、論理パス情報（例えばポート番号、ターゲットID及びLUN）、アドレス管理情報（例えば、ディスク型記憶装置300上の物理アドレスとLEDEVにおける論理アドレスとの対応関係）、記憶容量及びRAIDレベルを有する。なお、物理アドレスとは、例えば、ディスク型記憶装置300のID、ディスクヘッド番号、及びセクタ数を含んだアドレス情報である。論理アドレスとは、例えば、LUN（Logical Unit Number）、LEDEV番号、及び論理ブロックアドレスを含んだアドレス情報である。

【0035】

接続部150は、各チャネル制御部110A～110Dと、各ディスク制御部140A～140Dと、CM130と、SM120とを相互に接続するものである。チャネル制御部110A～110D、CM130、SM120及びディスク制御部140A～140D間でのデータやコマンドの授受は、接続部150を介することにより行われる。接続部150は、例えば、ユーザデータが通過する第1サブ接続部と、制御情報等が通過する第2接続部とを含んでいる。第1サブ接続部には、各チャネル制御部110A～110D、各ディスク制御部140A～140D及びCM130が接続され、第2サブ接続部には、各チャネル制御部110A～110D、各ディスク制御部140A～140D及びSM120が接続される。第1サブ接続部及び第2サブ接続部のうちの少なくとも第1サブ接続部は、例えば、高速スイッチングによりデータ伝送を行う超高速クロスバススイッチ等の高速バスである。

【0036】

複数のディスク型記憶装置300の各々は、例えば、ハードディスクドライブ或いは半導体メモリ装置等である。複数のディスク型記憶装置300のうちの2以上の所定数のディスク型記憶装置300によって、RAIDグループ2が構成されている。RAIDグループ2は、例えば、パリティグループ又はエラーコレクショングループとも呼ばれることがあり、RAIDの原理に従ったディスク型記憶装置300のグループである。同じRAIDグループ2に属する2以上のディスク型記憶装置300は、例えば異なるマザーボード上に搭載され、一つのディスク型記憶装置300が故障しても、残りの他のディスク型記憶装置300のデータを用いて、その故障したディスク型記憶装置300のデータを復元できるように構成されている。このRAIDグループ2の提供する物理的な記憶領域上に、論理的な記憶デバイスである複数のLEDEV（Logical Device）が設定され、複数のLEDEVのうちの1以上のLEDEVが、LUN（Logical Unit Number）を持った1つのLU（Logical Unit）310として、記憶制御装置100からホスト端末200A～200Dに提供される。各LU310は、例えば、プライマリLU（データコピー元LU）として、セカンダリLU（データコピー先LU）である別LU310とペアになる場合があり、その場合、そのLU310内の全部又は一部のデータ（例えば更新前のデータ）が、別LU310にコピーされることがある。また、各LU310は、例えば、セカンダリLUとして、プライマリLUである他のLU310とペアになる場合があり、その場合、別LU310内の全部又は一部のデータ（例えば更新前のデータ）が、そのLU310にコピーされることがある。

【0037】

SVP160は、ストレージシステム600を保守又は管理するためのコンピュータマシンである。SVP160は、例えば、内部LAN等の通信ネットワーク151を介して、記憶制御システム600の各構成要素（例えば、各チャネル制御部110A～110D及び各ディスク制御部140A～140D）から情報を収集することができる。具体的には、例えば、記憶制御システム600の各構成要素（例えばチャネル制御部又はディスク制御部）に搭載されているOS（オペレーティングシステム）、アプリケーションプログラム、ドライバソフトウェア等が、その構成要素で発生した障害発生に関する障害発生情報を出力するようになっていて、SVP160が、その障害発生情報を受信することがで

きる。SVP160が受ける情報としては、例えば、装置構成、電源アラーム、温度アラーム、入出力速度（例えば、記憶制御装置100が一定時間当たり受信したI/O要求の数）等がある。また、例えば、SVP160は、オペレータによる操作に応答して、例えば、ディスク型記憶装置300の設定や、LDEVの設定や、チャンネル制御部110A～110Dにおいて実行されるマイクロプログラムのインストール等を行うことができる。また、SVP160は、例えば、記憶制御システム600の動作状態の確認や故障部位の特定、チャンネル制御部110で実行されるオペレーティングシステムのインストール等の作業を行うこともできる。また、例えば、SVP160は、LANや電話回線等の通信ネットワークを介して外部保守センタ（図示せず）と接続されていて、その外部保守センタに、記憶制御システム600の各構成要素から受信した障害発生情報等を通知しても良い。また、SVP160は、記憶制御システム600に内蔵されている形態とすることもできるし、外付けされている形態とすることもできる。

【0038】

以上が、記憶制御システム600についての基本的な説明である。なお、この記憶制御システム600では、例えば、1つのチャンネル制御部と1つのディスク制御部とが1つのモジュールとして一体的に構成されて、その1つのモジュールで、チャンネル制御部及びディスク制御部の機能を発揮しても良い。また、例えば、SM120とCM130が、一体的に構成されていても良い。また、各チャンネル制御部毎に1つのLUNが割当てられても良いし、複数のチャンネル制御部に1つのLUNが割当てられても良い。また、記憶制御システム600には、別の記憶制御システムが接続されていても良い。その場合、例えば、記憶制御システム600が有するプライマリLUと、別の記憶制御システムが有するセカンダリLUとがペア状態にされて、記憶制御システム600に接続されているホスト端末200A～200Dが、記憶制御システム600を介して別の記憶制御システム内のセカンダリLUにアクセスしても良い。また、例えば、記憶制御システム600は、ファイルアクセス要求及びブロックアクセス要求の双方を受けるものであっても良いし、ファイルアクセス要求のみを受けるもの（例えばNAS）であっても良いし、ブロックアクセス要求のみを受けるものであっても良い。

【0039】

図3は、本実施形態に係る記憶制御サブシステム102の機能を示すブロック図である。なお、以下の説明では、説明を分かり易くするために、ホスト端末200A～200Dのうちホスト端末200Aを例に採り、且つ、チャンネル制御部110A～110Dのうちチャンネル制御部110Aを例に採る。

【0040】

ホスト端末200Aと記憶制御サブシステム102との間には、1又は複数の論理的な通信パス（以下、「論理パス」と言う）21A～21Dが形成されている。各論理パス21A～21Dは、例えば、記憶制御サブシステム102が有するポート（ホスト端末100Aが接続されるポート）の番号と、ターゲットIDと、LUNとに基づいて形成されたものである。

【0041】

1又は複数のRAIDグループ2上に用意される複数のLDEVには、例えば、通常LDEVという属性を持ったLDEV（以下、通常LDEV）61aと、プールLDEVという属性を持ったLDEV（以下、プールLDEV）61bとがある。各LDEVは、SVP160のオペレータの指示により、通常LDEV61aからプールLDEV61bになることもできるし、逆に、プールLDEV61bから通常LDEV61aになることもできる。

【0042】

通常LDEV61aは、ホスト端末110Aがアクセス可能なLDEVである。換言すれば、例えば、論理パス21Aが有するLUNがホスト端末110Aから指定された場合、そのLUNに対応付けられている2つの通常LDEV61aが、LU310Aとしてホスト端末200Aに提供される。また、例えば、論理パス21Bが有するLUNがホスト

端末 110A から指定された場合、その LUN に対応付けられている 1 つの通常 LDEV 61a が、LU310B としてホスト端末 200A に提供される。

【0043】

プール LDEV 61b は、LDEV プール 68 を構成するメンバであり、ホストがアクセス不可能な LDEV である。換言すれば、プール LDEV 61b は、ホスト端末 200A から指定され得る LUN に対応付けられておらず、LUN が指定されてもプール LDEV 61b それ自体はホスト端末 200A に提供されないようになっている。

【0044】

LDEV プール 68 の上位には、ホスト端末 200A に対して提供される 1 又は複数の仮想 LU310C が存在する。仮想 LU310C は、ホスト端末 200A に提供される LU であるが、他の LU310A、310B と違い、RAID グループ 2 上に物理的なデータ格納領域を持たない仮想的な LU である。具体的に言えば、仮想 LU310C は、仮想 LDEV 61c から構成されるが、その仮想 LDEV 61c は、通常 LDEV 61a やプール LDEV 61b と異なり、RAID グループ 2 上に物理的なデータ格納領域を持たない仮想的な LDEV である。仮想 LDEV 61c が有する複数の仮想アドレスの各々は、ダイナミックマップテーブル（以下、「DMT」と略記）64 を介して、動的に、プール LDEV 68 が有する複数の論理アドレスから選択された論理アドレスに対応付けられたり、その対応付けが解除されてその論理アドレスが開放されたりする。

【0045】

図 4 は、仮想 LDEV 61c、LDEV プール 68、及び DMT 64 の構成例を示す。

【0046】

仮想 LDEV 61c は、例えば、一定サイズ（例えば 64 キロバイト）を有する複数の仮想チャンク 410c、410c、…から成っている。各仮想チャンク 410c は、所定個数（例えば 128 個）の論理ブロック（例えば 1 個が 512 バイト）から構成されている。各仮想チャンク 410c には、先頭論理ブロックアドレス（以下、「仮想先頭 LBA」と略記）が存在し、仮想先頭 LBA から、どの仮想チャンク 410c であるかを特定することができるようになっている。

【0047】

各 LDEV プール 68 は、1 以上の LDEV 61b の集合であり、例えば、一定サイズ（例えば 64 キロバイト）を有する複数の論理チャンク 410b、410b、…から成っている。各論理チャンク 61b は、所定個数（例えば 128 個）の論理ブロック（例えば 1 個が 512 バイト）から構成されている。各論理チャンク 410b には、先頭論理ブロックアドレス（以下、「論理先頭 LBA」と略記）が存在し、論理先頭 LBA から、どの論理チャンク 410c であるかを特定することができるようになっている。なお、例えば、アクセスする論理先頭 LBA が決まれば、対応するチャンクの番号が決まる。具体的には、例えば、論理チャンクサイズを 6.4 KB (= 128 ブロック) とすると、論理チャンク番号 n は、論理ブロック番号 m / 128 の整数部となる（例えば、論理ブロック番号が 0 ~ 127 の論理ブロックを有する論理チャンク番号はゼロであり、論理ブロック番号が 128 ~ 255 の論理ブロックを有する論理チャンクの番号は 1 である）。

【0048】

DMT 64 は、外部端末（例えば SVP 160）から記憶制御サブシステム 102 内の記憶領域（例えば SM 120）に登録される。DMT 64 は、各仮想 LDEV 61c 毎に用意され、その仮想 LDEV 61c の各仮想チャンク 410c と、1 以上の LDEV プール 68 の各々における各論理チャンク 410b とを対応付けるためのテーブルである。具体的には、例えば、DMT 64 には、各仮想先頭 LBA 毎に、その仮想先頭 LBA に対応付けられた LDEV プール 68 の識別情報（例えば LDEV プール番号）と、その LDEV プール 68 における論理先頭 LBA とが記述される。この DMT 64 の記述内容は、所定のタイミング、例えば、チャンネル制御部 110A が I/O 要求を受けた場合に、チャンネル制御部 110A によって更新される。チャンネル制御部 110A は、プールマネジメントテーブル（以下、「PMT」と略記）63 を参照して、DMT 64 の記述内容の更新を更

新する。

【0049】

図5は、PMT63の構成例を示す。

【0050】

PMT63は、LDEVプール68毎に存在するものであり、外部端末（例えばSVP160）から記憶制御サブシステム102内の記憶領域（例えばSM120）に登録される。LDEVプール68の各論理チャック401b毎に、エントリ番号が割り当てられる。各PMT63には、それに対応したLDEVプール68における初めの空きエントリの番号（換言すれば、論理先頭LBAが最も若いチャック番号）を筆頭に、キュー形式で、他の空きエントリの番号が書かれる。

【0051】

以下、仮想LDEV61cとLDEVプール68とを動的に対応付けるチャンネル制御部110Aの処理流れについて説明する。

【0052】

ホスト端末200Aは、通常のLU310Aにアクセスする場合と同様の方法で、仮想LU310Cに対するI/O要求を出力する。チャンネル制御部110Aは、例えば、ホスト端末110AからそのI/O要求を受けた場合、そのI/O要求から割り出された仮想先頭LBAと、LDEVプール68の論理先頭LBAとの対応付けを行う。具体的には、例えば、チャンネル制御部110Aは、I/O要求を受けた仮想LU310Cが有する仮想LDEV61cに対応したDMT64を参照し、上記割り出された仮想先頭LBAに対応したLDEVプール番号を取得する。次に、チャンネル制御部110Aは、取得したLDEVプール番号に対応したPMT63から、そのPMT63に書かれている最初の空きエントリ番号に対応した論理先頭LBAを取得し、その論理先頭LBAを、上記仮想先頭LBAに対応した場所（DMT64上の場所）に登録する。それにより、上記仮想先頭LBAに論理先頭LBAが対応付けられ、その論理先頭LBAを持つ論理チャックからデータが読み出されたり、その論理チャックにデータが書き込まれたりすることができるようになる。

【0053】

また、チャンネル制御部110Aは、例えば、上記I/O要求に基づく処理の終了後に、DMT64に書かれた論理先頭LBAを開放することができる。具体的には、例えば、チャンネル制御部110Aは、DMT64に登録した上記論理先頭LBAをDMT64から削除し、PMT63に、その論理先頭LBA及びそれに対応するエントリ番号をPMT64に格納する。

【0054】

以上のような流れで、仮想LU310C（仮想LDEV61c）とLDEVプール68との動的な対応付けやその解除が行われる。このように、どのホスト端末からアクセスされるかが予め割当てられていないプールLDEV61b、61b、…から成るLDEVプール68を用意し、そのLDEVプール68における記憶領域をフレキシブルに仮想LU310C（仮想LDEV61c）に対応付けることにより、空いている記憶領域を有効に活用することができる。

【0055】

図6は、この実施形態に係るLU-LDEV管理テーブル162bの構成例を示す。

【0056】

LU-LDEV管理テーブル162bは、記憶制御サブシステム102内の記憶領域（例えばSM120）に登録されるものである。LU-LDEV管理テーブル162bには、例えば、記憶制御システム600が備える各ポート（ホスト端末が接続されるポート）毎に、ポート番号と、そのポート番号に属する1以上のターゲットIDと、各ターゲットIDに属する1以上のLUNと、各LUNに属する1以上のLDEVに関するLDEV情報が記録されている。各LDEVのLDEV情報としては、例えば、LDEV番号、記憶容量、RAIDレベル、上述したアドレス管理情報、LDEV属性、DMT-ID、LD

EV プール番号; 状態及び記憶容量通知済みホスト ID がある。

【0057】

LDEV 属性としては、例えば、上述したプール LDEV、通常 LDEV 及び仮想 LDEV というものがある。

【0058】

「プール LDEV」という LDEV 属性を持つ LDEV (前述したプール LDEV 61b) は、ホスト端末にアクセスされないようになっているため、プール LDEV 61b には、論理パス情報 (例えば LUN) は対応付けられていない。また、プール LDEV 61b には、どの LDEV プール 68 のメンバであるかが識別されるように LDEV プール番号が対応付けられている。

【0059】

「仮想 LDEV」という LDEV 属性を有する LDEV (前述した仮想 LDEV 61c) には、その LDEV 61c に対応する DMT 64 の ID (例えばゼロ) が対応付けられる。この DMT-ID を参照することにより、どの仮想 LDEV 61c (仮想 LU 310C) にアクセスされた場合にはどの DMT を参照すれば良いかがわかるようになっている。また、仮想 LDEV 61c には、どの LDEV プール 68 のメンバであるかが識別されるように LDEV プール番号が対応付けられている。

【0060】

「通常 LDEV」という LDEV 属性を有する LDEV (前述した通常 LDEV 61a) には、DMT-ID 及び LDEV プール番号は対応付けられていない。

【0061】

LDEV の状態としては、例えば、アクセス可能な状態であることを表す「Ready」 (例えば実装状態) と、アクセス不可能な状態であることを表す「Not Ready」 (例えば未実装状態) とがある。

【0062】

各 LDEV の記憶容量通知済みホスト ID とは、LDEV が属する LU の記憶容量の通知を受けたホスト端末の ID (例えば、MAC アドレス又は WWN (World Wide Name)) である。これを参照することにより、どの LU の記憶容量がどのホスト端末に通知されたかが分かるようになっている。

【0063】

図 7 は、チャンネル制御部 110A の構成例を示す。

【0064】

チャンネル制御部 110A は、ハードウェア要素が一体的にユニット化されたボードで構成される。チャンネル制御部 110A は、CHN (Channel adapter Nas) 又は NAS ブレードと呼ばれることがある。チャンネル制御部 110A は、図示しないボード接続用コネクタを備えており、そのボード接続用コネクタが記憶制御装置 100 の所定コネクタに嵌合することにより、チャンネル制御部 110A は記憶制御装置 100 と電氣的に接続される。チャンネル制御部 110A は、例えば、ホストインターフェース部 (以下、「ホスト I/F」と記載) 711、SVP インターフェース部 (以下、「SVP I/F」と記載) 51、ブリッジ LSI (Large-Scale Integration) 781、メモリコントローラ 741、NAS プロセッサ 112、ホスト側メモリ 113、1 又は複数の入出力制御部 771 及びデータ転送 LSI 782 を備える。

【0065】

ホスト I/F 711 は、ホスト端末 200A 及び 200B との間で通信を行うための通信インタフェースであり、例えば、TCP/IP プロトコルに従ってホスト端末 200A 及び 200B から送信されたファイルアクセス要求を受信する。

【0066】

SVP I/F 51 は、内部 LAN 150 等の通信ネットワークを介して SVP 160 に接続され、且つ、後述の CHP 119 に接続される。SVP I/F 51 は、SVP 160 と CHP 119 との間の通信を制御するための通信インタフェース (例えば LAN コント

ローラ)である。

【0067】

ブリッジLSI 781は、例えば、ホストI/F 711、メモリコントローラ741及びデータ転送LSI 782の間の通信を可能にするためのLSIである。

【0068】

メモリコントローラ505は、ブリッジLSI 782、NASプロセッサ112及びホスト側メモリ113の間の通信を可能にするためのLSIである。

【0069】

NASプロセッサ112は、例えばCPUであり、NFS (Network File System) を用いて、記憶制御サブシステム600をNASとして機能させるための制御を行う。例えば、NASプロセッサ112は、NFS等により、ファイル名と論理ブロックアドレスとの対応関係を把握しており、チャンネル制御部110Aが受けたファイルアクセス要求に含まれているファイル名及び上記対応関係を基に、そのファイルアクセス要求をブロックアクセス要求に変換して、CHP 121に出力する。

【0070】

ホスト側メモリ113には、例えば、様々なプログラムやデータが記憶される。具体的には、例えば、ホスト側メモリ113には、後述するファイルメタデータテーブル、ロックテーブル、及びNASマネージャ等のデータやプログラムが記憶される。

【0071】

各入出力制御部771は、ディスク制御部140A~140D、CM130、SM120及びSVP160との間でデータやコマンドの授受を行うもの（例えばマイクロプロセッサユニット）である。各入出力制御部771は、CHP（チャンネルプロセッサ）121や、CHPメモリ119を備える。

【0072】

CHP 121は、例えば、マイクロプロセッサであり、チャンネル制御部110A全体の制御を司ると共に、ディスク制御部140A~140Dや、ホスト端末200A及び200Bや、SVP160との間の通信を制御する。CHP 121は、CHPメモリ121（又はホスト側メモリ113）に格納された各種コンピュータプログラムを実行することにより、チャンネル制御部110Aとしての機能を発揮する。

【0073】

CHPメモリ119は、揮発性又は不揮発性のメモリ（例えばNVRAM (Non Volatile RAM)）であり、例えば、CHP 121の制御を司るコンピュータプログラムを格納する。CHPメモリ119に記憶されるプログラムの内容は、例えば、SVP160や、後述するNASマネージャ806からの指示により書き込みや書き換えを行うことができる。

【0074】

データ転送LSI 501は、データ転送の制御を行うためのLSIである。具体的には、例えば、データ転送LSI 501は、CMデータ転送回路710及びSMデータ転送回路740を備える。CMデータ転送回路710は、CM130に接続され、ユーザデータの入出力を行う。SMデータ転送回路740は、SM120に接続され、制御情報の入出力を行う。

【0075】

上述したホスト側メモリ113（又はCHPメモリ121）には、例えば、リードキャパシティ処理プログラム715及びLU定義処理プログラム716が格納される。

【0076】

リードキャパシティ処理プログラム715は、ホスト端末200Aからリードキャパシティコマンドを受信した場合に、そのリードキャパシティコマンドが受けたポート番号、ターゲットID及びLUNに対応したLUの記憶容量をLU-LDEV管理テーブル162bから取得し、その記憶容量を、リードキャパシティコマンド発行元のホスト端末200Aに通知する。

【0077】

LU定義処理プログラム716は、SVP160からのアクセスに応答して、LU設定画面をSVP160に提供し、そのLU設定画面に入力された情報に基づいて、LU-LDEV管理テーブル162bを更新する。

【0078】

なお、リードキャパシティ処理プログラム715及びLU定義処理プログラム716の少なくとも一方は、必ずしもチャンネル制御部110Aに格納されている必要は無く、例えばディスク制御部140A～140Dに搭載されているメモリ等別の場所に格納されていても良い。

【0079】

図8は、LUが設定される場合に行なわれる記憶制御サブシステム102での処理流れを示す。

【0080】

この図に示す処理は、例えば、チャンネル制御部110Aが実行する。

【0081】

チャンネル制御部110Aは、SVP160（又はそれに接続された外部端末）から、LDEVプールの作成依頼を受ける。チャンネル制御部110Aは、その依頼に応答して、LDEVプール編集画面552をSVP160に提供する。この画面552には、例えば、作成対象のLDEVプール番号の入力欄と、そのLDEVプールに対して追加、除去又は編集するためのLDEV番号の入力欄とがある。また、例えば、その画面552には、LDEV番号の入力に入力されたLDEV番号を追加、除去又は編集のいずれを行うかの選択を受け付けるツールが設けられている。

【0082】

チャンネル制御部110Aは、そのLDEVプール編集画面552に入力された情報に従って、LDEVプールの作成（又は編集）を行う（換言すれば、LU-LDEV管理テーブル162bを更新する）（ステップS300）。

【0083】

その後、チャンネル制御部110Aは、仮想LUと通常LUのいずれを定義するかを選択をSVP160から受け、仮想LUが選択された場合、仮想LU設定画面551をSVP160に提供する。その画面551には、その仮想LUに対応付けるポート番号、ターゲットID、LUN、LDEV番号及び仮想記憶容量の入力欄がある。

【0084】

チャンネル制御部110Aは、その仮想LU設定画面551に入力された情報に従って、仮想LUを定義する（換言すれば、「仮想LDEV」というLDEV属性を持った新たな仮想LDEVに関する情報をLU-LDEV管理テーブル162bに登録する）（S310）。

【0085】

また、チャンネル制御部110Aは、S310で定義した仮想LUがどのLDEVプールを使用するかを指定を受け（例えば、そのLDEVプールの番号の入力を受け）、指定されたLDEVプールを仮想LU（仮想LDEV）に対応付ける（換言すれば、入力されたLDEVプール番号をLU-LDEV管理テーブル162bに登録する）（S320）。

【0086】

以上の流れにより、仮想LU（仮想LDEV）の定義付けが完了する。なお、ここで入力された仮想記憶容量は、ホスト端末200Aからのリードキャパシティコマンドに応答して、そのホスト端末200Aに通知される。

【0087】

図9は、ホスト端末200Aの処理フローの一例を示す。

【0088】

ホスト端末200Aは、例えば、再起動処理を実行する場合、又は、ホスト端末200Aに接続されている全ての記憶制御システム600を調査する再スキャン処理を実行する

場合、以下の処理フローを実行する。

【0089】

例えば、ホスト端末200Aは、第1のコマンド（例えばInquiryコマンド）を記憶制御システム600に発行して、その記憶制御システム600に関する制御システム情報（例えば、その記憶制御システム600のベンダ名やモデル名）をその記憶制御システム600から受信する（S201）。

【0090】

次に、ホスト端末200Aは、第2のコマンド（例えばデバイスディスカバリーコマンド又はリードキャパシティコマンド）を記憶制御システム600に発行して、その記憶制御システム600がホスト端末200Aに提供する全てのLUに関するLU情報を、その記憶制御システム600から受信する（S202）。受信するLU情報には、ホスト端末200Aがアクセス可能な各LU毎に、例えば、少なくとも記憶容量情報が含まれており、その他、そのLUのパス（例えば、ポート番号、ターゲットID及びLUN）や、そのLUに属するLDEVを持ったディスク型記憶装置のベンダ等も含まれていても良い。

【0091】

ホスト端末200Aは、S201で受信した制御システム情報や、S202で受信したLU情報を、ホスト端末200Aの記憶装置（例えばハードディスクドライブ）に格納する（S203）。例えば、ホスト端末200Aは、S202で受信したLU情報に基づいて、各記憶制御システム600毎に、図10に例示するような記憶制御システム管理テーブル791を作成して記憶する。記憶制御システム管理テーブル791には、各LU毎に、ポート番号、ターゲットID、LUN、ベンダ、モデル及び記憶容量等が登録される。なお、LUが仮想LUの場合、S202で受信する記憶容量は、仮想LU設定画面551で入力された値（すなわち、オペレータ任意の値）である。

【0092】

ホスト端末200Aは、全ての記憶制御システム600に対して、S201～S203処理を実行する（S304でN）。

【0093】

図11は、チャネル制御部110Aが、ホスト端末200から第2コマンドとしてリードキャパシティコマンドを受けた場合に行なう処理フローの一例を示す。

【0094】

チャネル制御部110Aは、ホスト端末200Aからリードキャパシティコマンドを受けた場合、そのリードキャパシティコマンドを受けたポート番号、ターゲットID及びLUNに対応した記憶容量をLU-LDEV管理テーブル162bから取得し（S211）、その記憶容量をホスト端末200Aに通知する（S212）。また、チャネル制御部110Aは、通知先ホスト端末200AのホストIDを、LU-LDEV管理テーブル162bに書き込む。

【0095】

以上の説明では、説明を分かり易くするために、チャネル制御部110A及びホスト端末200Aを例に採り説明したが、上記説明は、その他のチャネル制御部110B～110D、ディスク制御部140A～140D、及びその他のホスト端末120B～120Dにも当てはまるものである。

【0096】

ところで、本実施形態では、更に、例えば以下の特徴的な処理が行われる。以下、上述の説明と重複するかもしれないが、幾つかの特徴的な処理について説明する。

【0097】

(1) 第1の特徴的な処理

図12は、第1の特徴的な処理に関わる処理流れを示す。

【0098】

記憶制御サブシステム102の記憶制御装置100（例えば、チャネル制御部又はディスク制御部）は、仮想LU310Cの仮想記憶容量値を所定記憶領域（例えばSM120

)に登録した場合に、仮想LU310Cを実装状態にする(換言すれば、仮想LU310Cをホスト端末に対してアクセス可能にする)(S1)。

【0099】

その後、記憶制御装置100は、SVP160(又はその他の外部端末)から、仮想記憶容量値の変更依頼を受けた場合(S2)、S1で登録した仮想記憶容量値をホスト端末に通知済みか否かを判別する(S3)。具体的には、例えば、記憶制御装置100は、LU-DEV管理テーブル162bを参照し、変更依頼対象の記憶容量値に対応した通知済みホストIDが登録されているか否かを判別する。

【0100】

S3の結果、肯定的な判別結果が得られた場合(S3でY)、記憶制御装置100は、SVP160からの記憶容量値変更依頼を拒否する(S4)。すなわち、記憶制御装置100は、一度でも過去にホスト端末に仮想記憶容量値を通知していれば、仮想記憶容量値の変更を禁止する。

【0101】

一方、S3の結果、否定的な判別結果が得られた場合(S3でN)、記憶制御装置100は、SVP160からの記憶容量値変更依頼を受けると共に、SVP160から変更後の仮想記憶容量値の入力を受け、それを、所定記憶領域における旧い仮想記憶容量値に上書きする(S5)。

【0102】

以上の処理により、一度でも過去にホスト端末に仮想記憶容量値が通知されて、そのホスト端末でその仮想記憶容量値が記憶されている可能性があれば、同一の仮想LUについて別の仮想記憶容量値がそのホスト端末に通知されて、そのホスト端末を混乱させてしまわないようなことがないように制御される。

【0103】

(2) 第2の特徴的な処理

記憶制御装置100は、オペレータに任意に設定された仮想記憶容量値を所定記憶領域に登録し、ホスト端末から第2コマンドを受信したら、そのコマンドに応じて、上記オペレータに設定された任意値をホスト端末に通知する。

【0104】

(3) 第3の特徴的な処理

図13は、第3の特徴的な処理に関わる構成及び処理流れを示す。

【0105】

記憶制御装置100は、2つのLUから成るLUペアを形成し、一方のLUをプライマリLU310P、他方のLUをセカンダリLU310Sとして、プライマリLU310P内のデータをセカンダリLU310Sにコピーするスナップショットを行うようになっている。

【0106】

具体的には、例えば、記憶制御装置100は、外部端末(例えばSVP160)からの要求に手動で、又は、ライト要求を示すI/O要求を受領した場合(S11)に自動で、通常LU310AをプライマリLU310Pとし、仮想LU310CをセカンダリLU310SとしたLUペアを形成する。

【0107】

その後、記憶制御装置100は、受領したライト要求から、通常LU310Aにおける書き込み先チャック(例えば先頭LBA)410Pを特定する(S12)。そして、記憶制御装置100は、特定されたチャック410P内の一部のデータ(斜線で示した部分)しか更新されなくても、そのチャック410P内の全てのデータを読み出し(S13)、そのデータを、仮想LU310Cに書き込む(S14)。具体的には、記憶制御装置100は、前述したように、仮想LU310C内の仮想チャック410Sと、LDEVプール68内の論理チャックとを動的に対応付けて、そのデータを、仮想チャック410Sに対応付けられたLDEVチャックに書き込む。

【0108】

このようにして、スナップショットを作成した後、記憶制御装置100は、上記特定されたチャンク410Pに、S11で受領したライト要求に含まれているデータを書き込む(S15)。

【0109】

なお、上記流れにおいて、記憶制御装置100は、LUペアを形成しようとする場合、LU-LDEV管理テーブル162bを参照し、プライマリLU310PとセカンダリLU310Sとの記憶容量値が一致した場合にのみLUペアを形成し、それらが一致しない場合、LUペアの形成を拒否しても良い。

【0110】

プライマリLU310PとセカンダリLU310Sとの記憶容量値が同一でないとLUペアが形成されないようになっている場合、プライマリLU310Pの記憶容量が大容量であると、セカンダリLU310Sの記憶容量も大容量である必要がある。しかし、スナップショット作成では、プライマリLU310PからセカンダリLU310Sにコピーするデータサイズが、プライマリLU310Pの容量よりも小さい場合がある。それにもかかわらず、大容量の通常LU310AをセカンダリLU310Sとしてしまうと、大量の無駄な空き記憶領域が生じてしまうことが考えられる。

【0111】

しかし、以上のように、スナップショットの作成において、セカンダリLU310Sを仮想LU310CとしてLUペアを形成することにより、その仮想LU310Cに対応付けられ得るLDEVプール68の空き領域は動的に別の仮想LU310Cに対応付けられるようになっているので、無駄に空き記憶領域が生じてしまうことを防ぐことができる。

【0112】

(4) 第4の特徴的な処理

図14は、第4の特徴的な処理に関わる処理流れを示す。

【0113】

記憶制御装置100は、例えば、仮想記憶容量値を持たない仮想LU310Cを準備する。

【0114】

その後、記憶制御装置100は、LUペア形成コマンドを外部端末(例えばSVP160)から受信した場合(S31)、そのLUペア形成コマンドで指定されたセカンダリLU310Sである仮想LU310Cの仮想記憶容量値が所定記憶領域に登録されているか否か(例えば、LU-LDEV管理テーブル162bに記憶容量値の登録があるか否か)を判別する(S32)。

【0115】

S32の結果、肯定的な判別結果が得られれば(S32でY)、記憶制御装置100は、S31で受信したLUペア形成コマンドに従って、LUペアを形成する(S34)。

【0116】

一方、S32の結果、否定的な判別結果が得られれば(S32でN)、記憶制御装置100は、上記LUペア形成コマンドで指定されたプライマリLU310P(例えば通常LU310A又は仮想LU310C)の記憶容量値を仮想記憶容量値として所定記憶領域に登録する(S33)。その後、記憶制御装置100は、S34の処理を実行する。

【0117】

(5) 第5の特徴的な処理

図15は、第5の特徴的な処理に関わる処理流れを示す。

【0118】

記憶制御装置100は、例えば、仮想記憶容量値を持たない仮想LU310Cを準備する。また、記憶制御装置100は、各LUペアに関するLUペア情報(例えばLUペアを構成する各LUのLUN)を所定の記憶領域(例えばSM120)で管理している。

【0119】

記憶制御装置 1 0 0 は、そのような仮想 L U 3 1 0 C に対する I / O 要求を受信した場合 (S 4 1) 、上記 L U ペア情報を参照して、その仮想 L U のペア相手が存在するか否かを判別する (S 4 2) 。

【 0 1 2 0 】

S 4 2 の判別の結果、否定的な判別結果が得られた場合 (S 4 2 で N) 、記憶制御装置 1 0 0 は、L U が未実装である旨を、I / O 要求送信元のホスト端末に通知し (S 4 7) 、処理を終える。なお、その後、記憶制御装置 1 0 0 は、その仮想 L U の相手 L U が設定された後、所定のタイミングで (例えば、ペア相手 L U が設定された時点で)、その相手 L U の記憶容量値を仮想 L U の仮想記憶容量値として所定記憶領域に登録する。また、記憶制御装置 1 0 0 は、L U が未実装である旨の通知先となったホスト端末のホスト I D を記憶しておき、仮想記憶容量値に登録した時点で、その仮想記憶容量値を、そのホスト I D に対応したホスト端末に通知しても良い。

【 0 1 2 1 】

一方、S 4 2 の判別の結果、肯定的な判別結果が得られた場合 (S 4 2 で Y) 、記憶制御装置 1 0 0 は、仮想記憶容量値が所定記憶領域に登録されているか否かを判別する (S 4 3) 。

【 0 1 2 2 】

この S 4 3 の判別の結果、肯定的な結果が得られた場合 (S 4 3 で Y) 、記憶制御装置 1 0 0 は、S 4 1 で受信した I / O 要求に従う処理を実行する (S 4 5) 。

【 0 1 2 3 】

一方、S 4 3 の判別の結果、否定的な結果が得られた場合 (S 4 3 で N) 、記憶制御装置 1 0 0 は、仮想 L U の相手 L U の記憶容量値を仮想記憶容量値として所定記憶領域に登録すると共に、その仮想記憶容量値を、I / O 要求送信元のホスト端末に通知する (S 4 4) 。その後、記憶制御装置 1 0 0 は、S 4 5 の処理を実行する。

【 0 1 2 4 】

(6) 第 6 の特徴的な処理

図 1 6 は、第 6 の特徴的な処理に関わる処理流れを示す。

【 0 1 2 5 】

ホスト端末は、所定の処理、例えば、再起動処理を実行する場合、又は、そのホスト端末に接続されている全ての記憶制御システム 6 0 0 を調査する再スキャン処理を実行する場合、過去に記憶した情報 (例えば、記憶制御システム管理テーブルそれ自体) を消去するようになっている。

【 0 1 2 6 】

記憶制御装置 1 0 0 は、例えば、外部端末 (例えば S V P 1 6 0) から、仮想記憶容量値の更新依頼を受けた場合 (S 5 1) 、その仮想記憶容量値を持つ仮想 L U 3 1 0 C にアクセス可能な全てのホスト端末との接続を切断する (換言すれば、その仮想 L U 3 1 0 C を強制的にオフライン状態にする) (S 5 2) 。

【 0 1 2 7 】

そして、記憶制御装置 1 0 0 は、外部端末から新たな仮想記憶容量値の入力を受け、古い仮想記憶容量値にその新たな仮想記憶容量値を上書きする (S 5 3) 。

【 0 1 2 8 】

その後、記憶制御装置 1 0 0 は、接続が切断されたホスト端末に対して再起動処理又は再スキャン処理命令を出力する (S 5 4) 。

【 0 1 2 9 】

再起動処理又は再スキャン処理命令を受けたホスト端末は、その命令に応答して再起動処理又は再スキャン処理を実行する (S 5 5) 。その際、ホスト端末は、過去に記憶した情報 (例えば、記憶制御システム管理テーブルそれ自体) を消去する。

【 0 1 3 0 】

S 5 5 の処理の完了後、ホスト端末は、処理終了通知を記憶制御装置 1 0 0 に送信する (S 5 6) 。なお、この処理終了通知に加えて又は代えて、上述した第 1 コマンド (例え

ばInquiryコマンド)や、第2コマンド(例えばデバイスディスカバリーコマンド又はリードキャパシティコマンド)が送信されても良い。

【0131】

記憶制御装置100は、ホスト端末から処理終了通知を受けたら、S53で上書きした新たな仮想記憶容量値を通知する(S57)。

【0132】

ホスト端末は、記憶制御装置100から通知された新たな仮想記憶容量値を記憶する(S58)。

【0133】

以上の処理流れにおいて、記憶制御装置100は、S53とS54の処理の間に、旧い仮想記憶容量値をホスト端末に通知済みか否かを判断する処理を行っても良い。その場合、肯定的な判断結果が得られた場合には、記憶制御装置100は、S54の処理を実行し、否定的な判断結果が得られた場合には、S54の処理を実行せずにS57の処理を行っても良い。

【0134】

以上、本発明の実施形態を説明したが、これは本発明の説明のための例示であって、本発明の範囲をこの実施形態にのみ限定する趣旨ではない。本発明は、他の種々の形態でも実施することが可能である。

【0135】

例えば、本発明に従う記憶制御サブシステムの第1実施態様では、前記記憶制御サブシステムの保守のための処理を行う保守用端末が前記記憶制御部に接続されている場合において、前記記憶制御部は、前記保守用端末又は前記保守用端末に接続された外部端末から、新たな前記仮想記憶ユニットを準備することのユニット準備要求を受け、前記ユニット準備要求に応答して、少なくとも前記仮想記憶容量値の記入欄を持ったグラフィカルユーザインターフェースを前記保守用端末又は前記外部端末に提供し、前記記入欄に入力された仮想記憶容量値を、前記設定された仮想記憶容量値として前記メモリに記憶させる。

【0136】

例えば、本発明に従う記憶制御サブシステムの第2実施態様では、前記記憶制御サブシステムが、2つの記憶ユニットから成るユニットペアを形成し、一方の記憶ユニットをプライマリ記憶ユニット、他方の記憶ユニットをセカンダリ記憶ユニットとして、プライマリ記憶ユニット内のデータをセカンダリ記憶ユニットにコピーするスナップショットを行うようになっている場合、前記物理記憶デバイスには複数の前記論理記憶デバイスが設けられており、前記複数の論理記憶デバイスには、前記仮想記憶領域に対応付けられ得る論理記憶領域を持った2以上の第1論理記憶デバイスと、前記仮想記憶領域に対応付けられることのない論理記憶領域を持った1以上の第2論理記憶デバイスとが含まれており、前記1以上の第2論理記憶デバイスが、前記ホスト端末に接続される1つのリアル記憶ユニットを構成し、前記記憶制御部は、前記リアル記憶ユニットを前記プライマリ記憶ユニットとし、前記仮想記憶ユニットをセカンダリ記憶ユニットとしたユニットペアを形成して前記スナップショットを行う。

【0137】

例えば、本発明に従う記憶制御サブシステムの第3実施態様では、前記第2実施態様において、前記記憶制御部は、前記仮想記憶容量値を前記ホスト端末に通知していない場合に、前記仮想記憶ユニットと前記リアル記憶ユニットの前記ユニットペアを形成するならば、前記リアル記憶ユニットの記憶容量値と同じ値を前記仮想記憶ユニットの記憶容量値として前記ホスト端末に通知する。

【0138】

例えば、本発明に従う記憶制御サブシステムの第4実施態様では、前記第3実施態様において、前記記憶制御部は、前記仮想記憶ユニットの相手となる前記リアル記憶ユニットが見つからない場合に、前記ホスト端末から前記仮想記憶ユニットに対してリード要求又はライト要求を受けたならば、前記仮想記憶ユニットが未実装状態であると前記ホスト端

末に通知し、その後、前記相手となるリアル記憶ユニットが見つかったならば、前記リアル記憶ユニットの記憶容量値と同じ値を前記仮想記憶ユニットの記憶容量値として前記ホスト端末に通知する。

【0139】

例えば、本発明に従う記憶制御サブシステムの第5実施態様では、前記記憶制御部は、前記ホスト端末からリードキャパシティコマンドを受信した場合に、前記メモリに記憶されている仮想記憶容量値を前記ホスト端末に通知する。

【0140】

例えば、本発明に従う記憶制御サブシステムの第6実施態様では、前記記憶制御サブシステムの保守のための処理を行う保守用端末が前記記憶制御部に接続されており、且つ、前記ホスト端末が行う起動処理又は再スキャン処理で、前記ホスト端末が、前記記憶した仮想記憶容量値を消去するようになっている場合において、前記記憶制御部は、前記仮想記憶容量値を前記ホスト端末に通知した後に、前記保守用端末又は前記保守用端末に接続された外部端末から、前記通知した仮想記憶容量値の更新要求を受けたならば、前記ホスト端末と前記仮想記憶ユニットとが接続されていない間に、前記保守用端末又は前記外部端末から新たな仮想記憶容量値を受けて前記メモリに記憶させ、且つ、前記ホスト端末に前記起動処理又は前記再スキャン処理を行わせることで、前記ホストに記憶されている旧い前記仮想記憶容量値が消去されるようにした後、前記メモリに記憶させた新たな仮想記憶容量値を前記ホスト端末に通知する。

【図面の簡単な説明】

【0141】

【図1】 本発明の一実施形態に係る記憶制御システムの外觀の概略を示す。

【図2】 本発明の一実施形態に係る記憶システムの全体構成例を示す。

【図3】 本実施形態に係る記憶制御サブシステムの機能を示すブロック図。

【図4】 仮想LDEV 61c、LDEV プール 68、及びDMT 64の構成例を示す。

【図5】 PMT 63の構成例を示す。

【図6】 本実施形態に係るLU-LDEV管理テーブル 162bの構成例を示す。

【図7】 チャンネル制御部 110Aの構成例を示す。

【図8】 LUが設定される場合に行なわれる記憶制御サブシステム 102での処理流れを示す。

【図9】 ホスト端末 200Aの処理フローの一例を示す。

【図10】 記憶制御システム管理テーブル 791の一例を示す図。

【図11】 チャンネル制御部 110Aが、ホスト端末 200から第2コマンドとしてリードキャパシティコマンドを受けた場合に行なう処理フローの一例を示す。

【図12】 本実施形態の第1の特徴的な処理に関わる処理流れを示す。

【図13】 本実施形態の第3の特徴的な処理に関わる構成及び処理流れを示す。

【図14】 本実施形態の第4の特徴的な処理に関わる処理流れを示す。

【図15】 本実施形態の第5の特徴的な処理に関わる処理流れを示す。

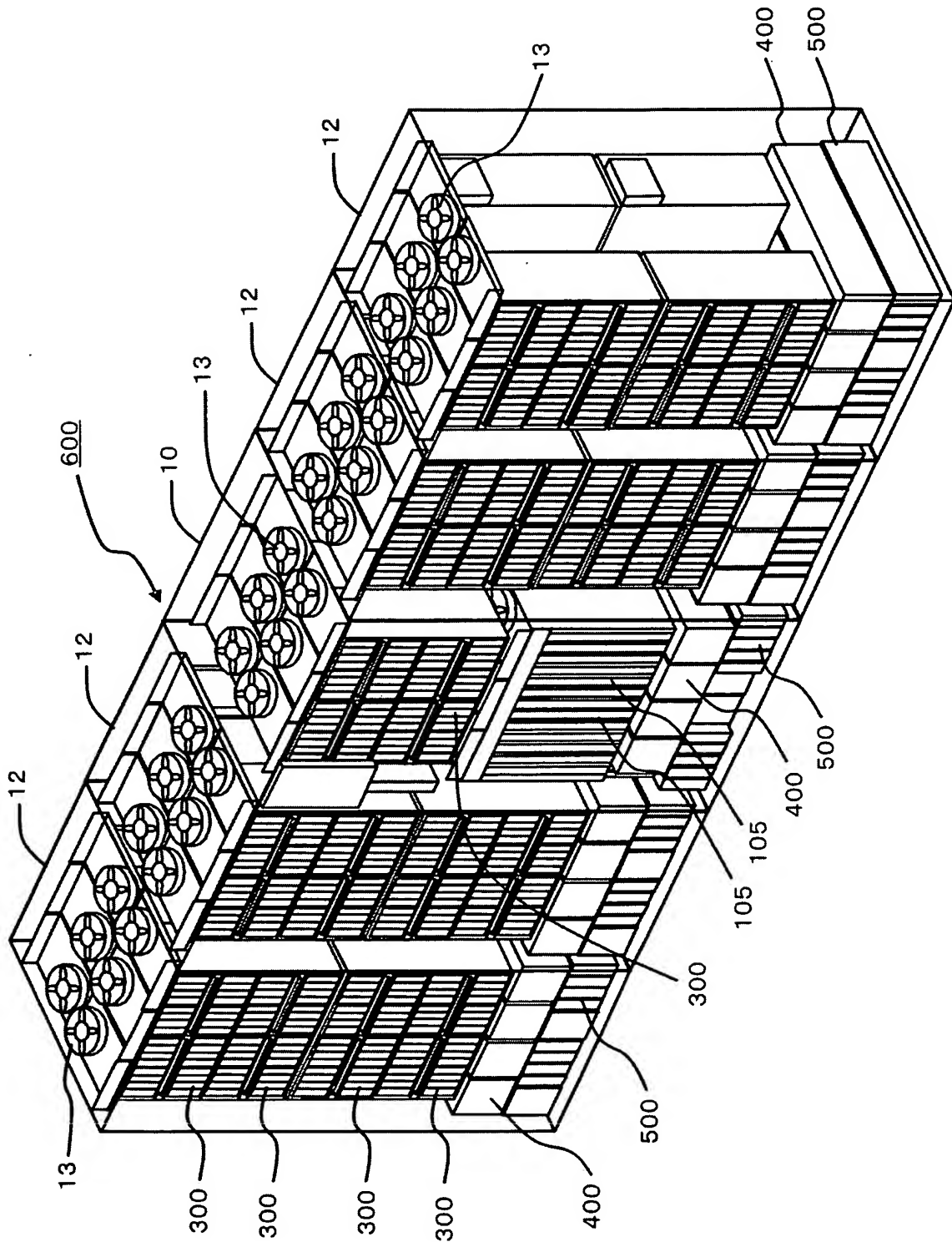
【図16】 本実施形態の第6の特徴的な処理に関わる処理流れを示す。

【符号の説明】

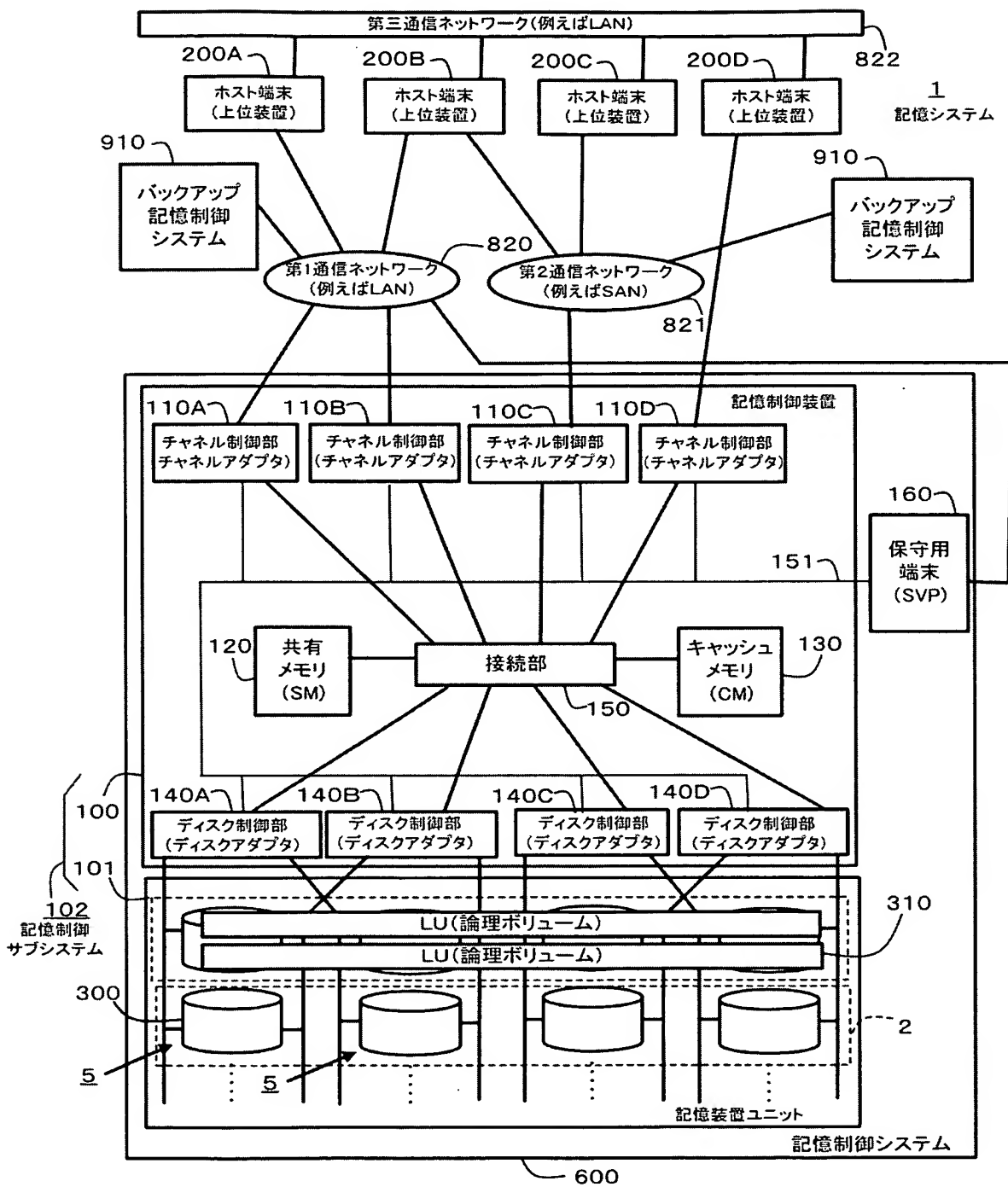
【0142】

1…記憶システム、2…RAIDグループ、10…基本筐体、12…増設筐体、100…記憶制御装置、101…記憶装置ユニット、102…記憶制御サブシステム、110A～110D…チャンネル制御部、120…共有メモリ、130…キャッシュメモリ、140A～140D…ディスク制御部、160…保守用端末、200A～200D…ホスト端末、300…ディスク型記憶装置、310…論理ユニット、600…記憶制御システム

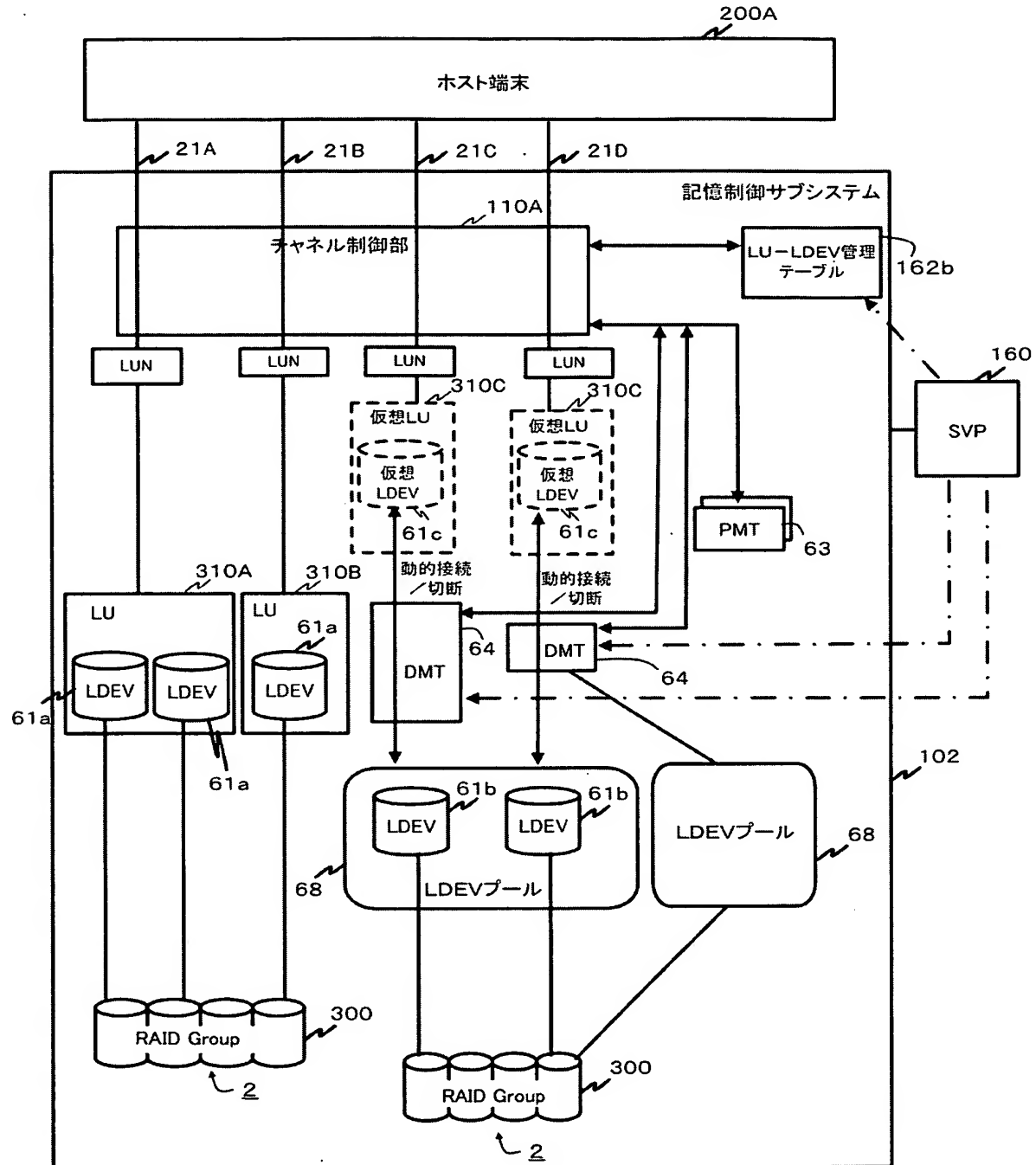
【書類名】図面
【図 1】



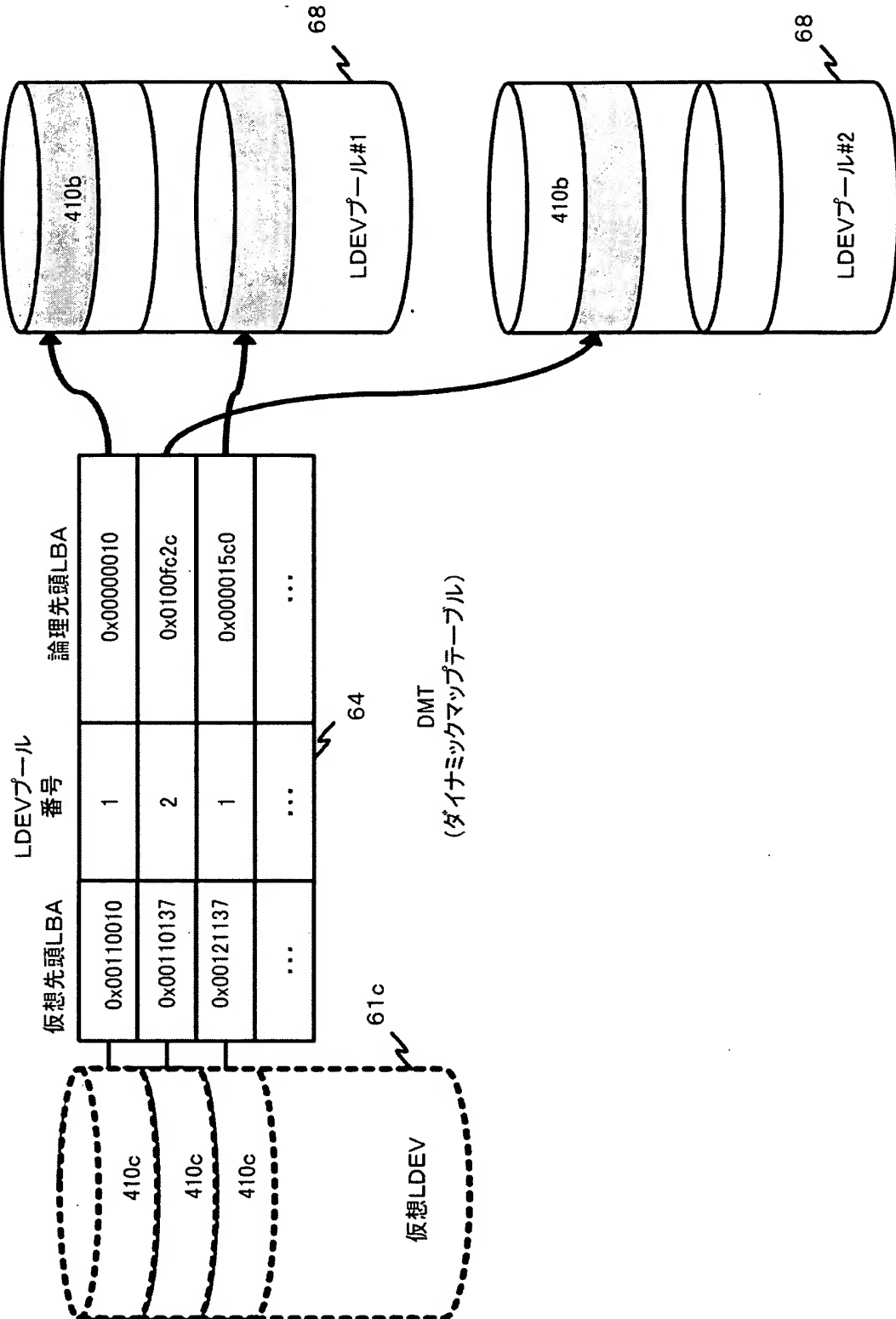
【図 2】



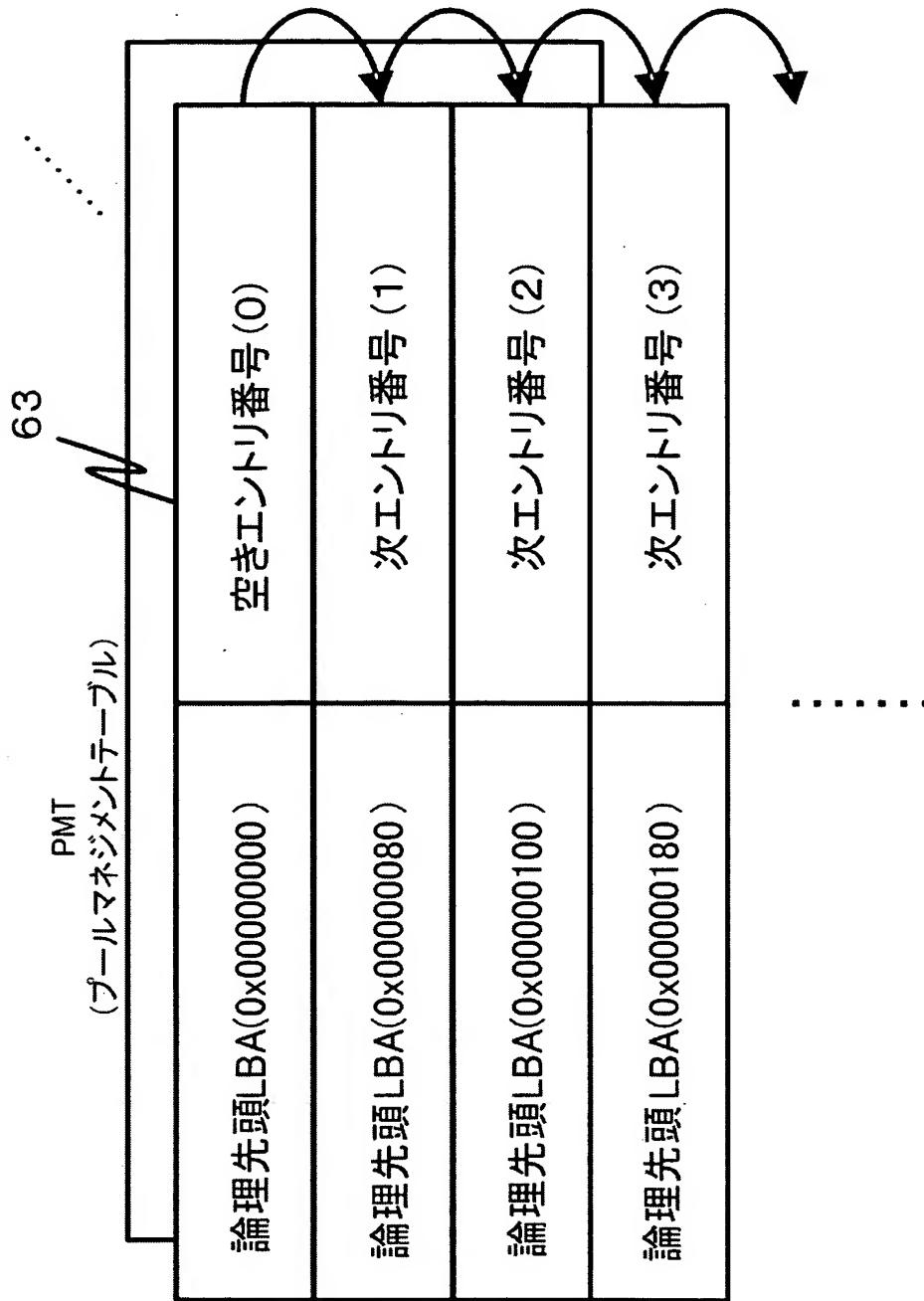
【図 3】



【図 4】



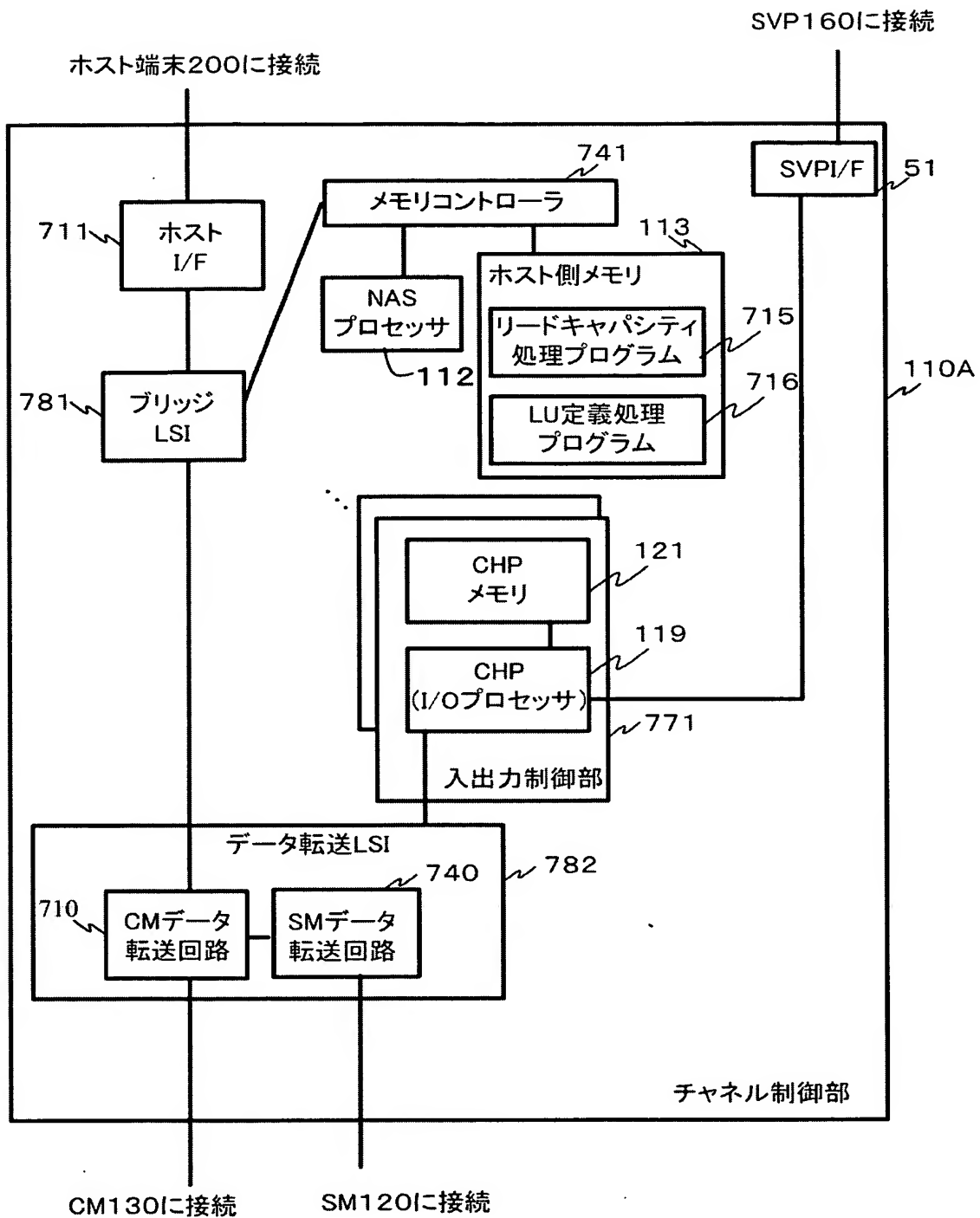
【図 5】



【図 6】

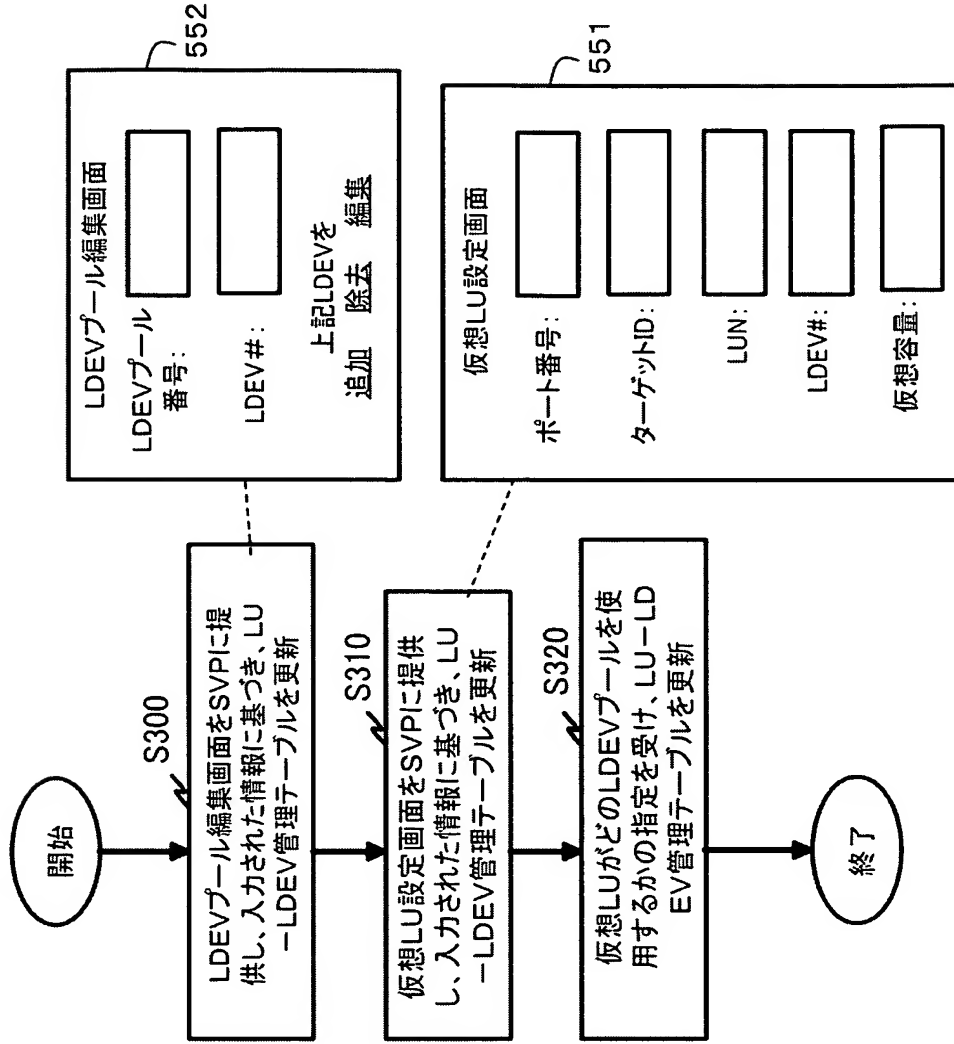
LU-LDEV管理テーブル											
ポート 番号	ターゲット ID	LUN	LDEV#	記憶容量	RAID レベル	アドレス管理 情報	LDEV属性	DMT-ID	LDEV プール番号	状態	通知済み ホストID
#1	-	-	LDEV#1	30GB	5	プールLDEV	-	#1	Ready
			LDEV#2	100GB	5	プールLDEV	-	#1	Ready
	#1	#002	LDEV#1	200GB	5	仮想LDEV	0	#1、#2	Ready
	#2	#2	#001	LDEV#1	50GB	1	通常LDEV	-	-
.....											
#2		#001	LDEV#1	50GB	1	通常LDEV	-	-	
										
.....											

【図 7】



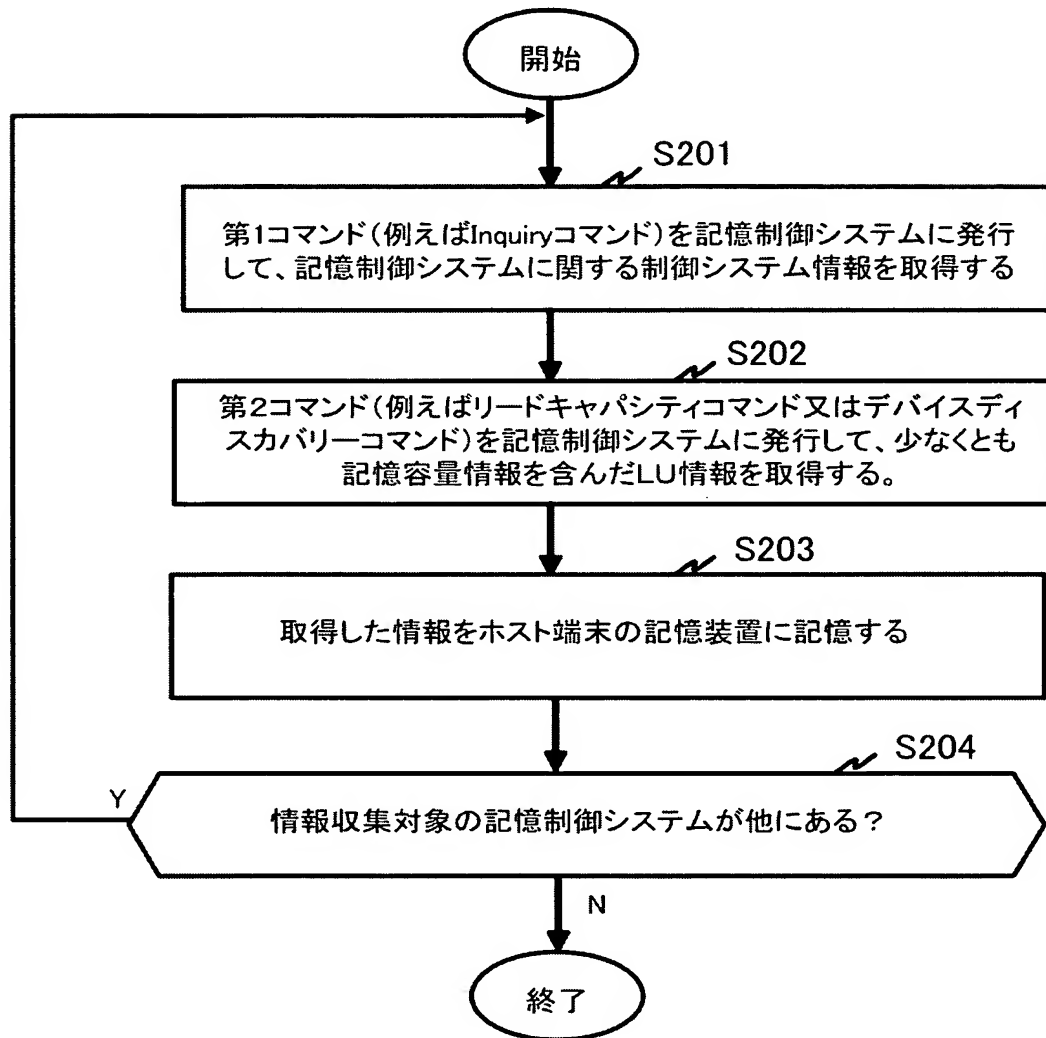
【図 8】

記憶制御サブシステム102
の動作フロー



【図9】

ホスト端末の動作フロー



【図 1 0】

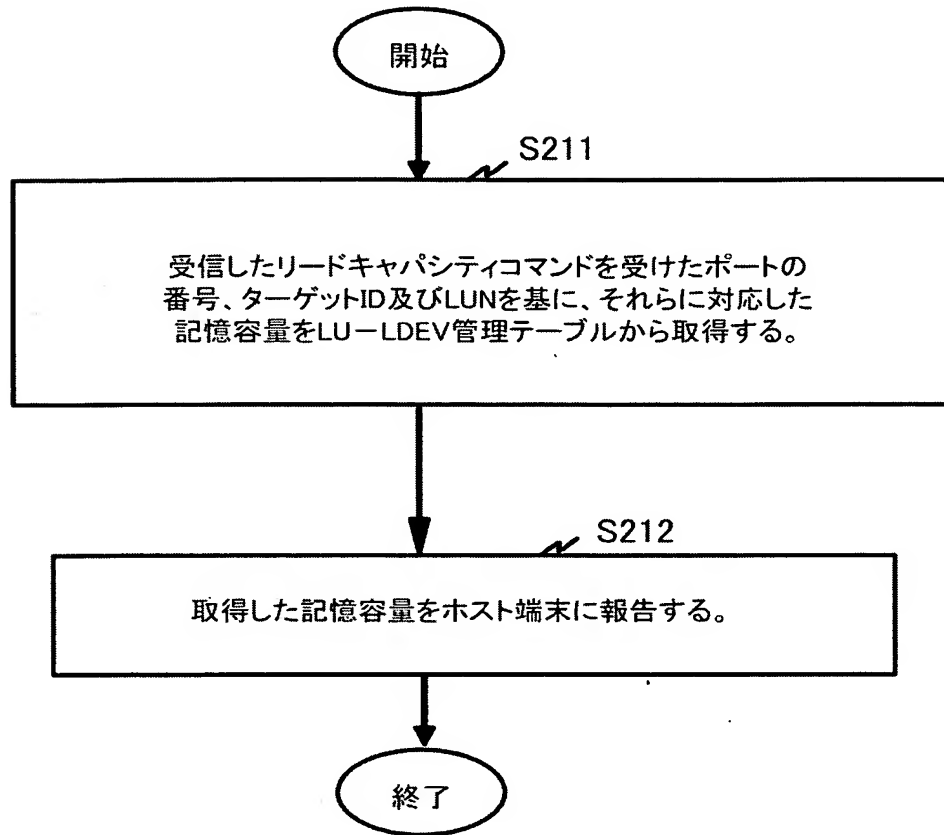
記憶制御システム
管理テーブル

791

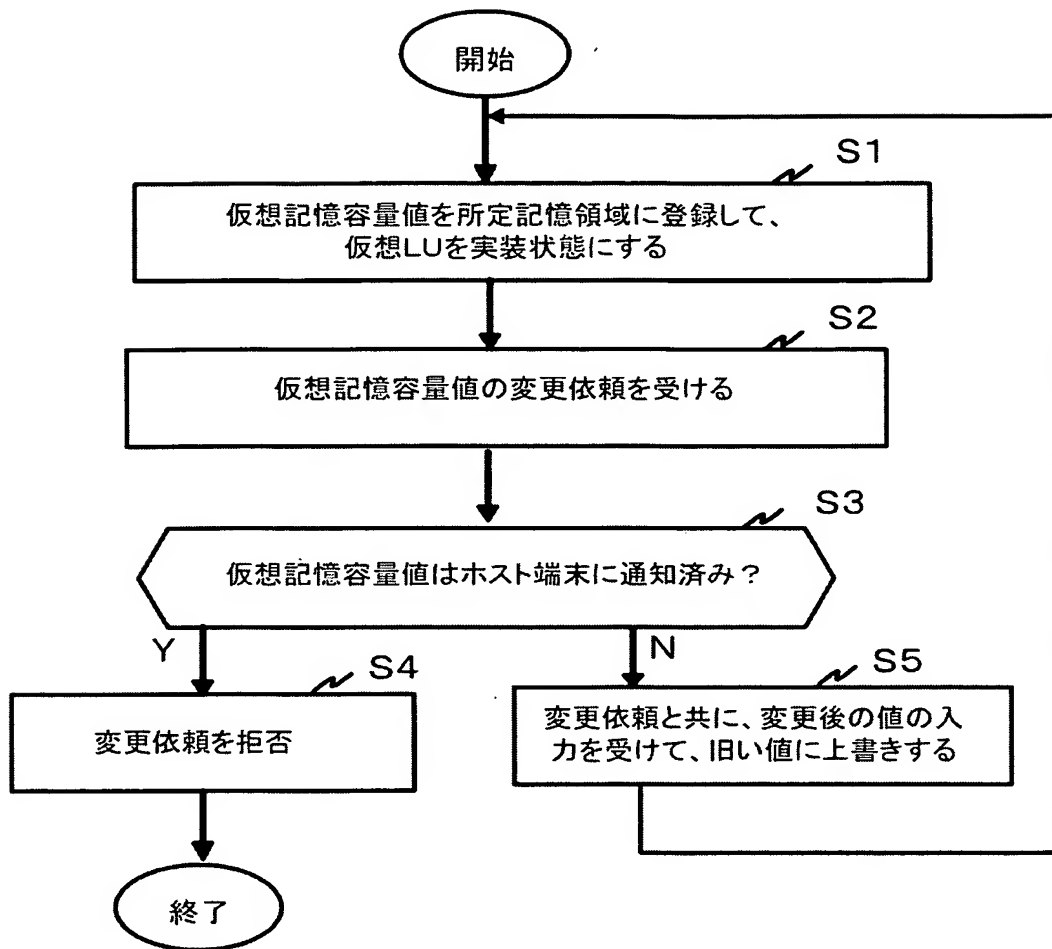
ポート番号	ターゲットID	LUN	ベンダ	モデル	記憶容量
0	0	0	H社	OPEN-3	1000GB
1	1	1	H社	OPEN-9	2000GB
.....

【図 11】

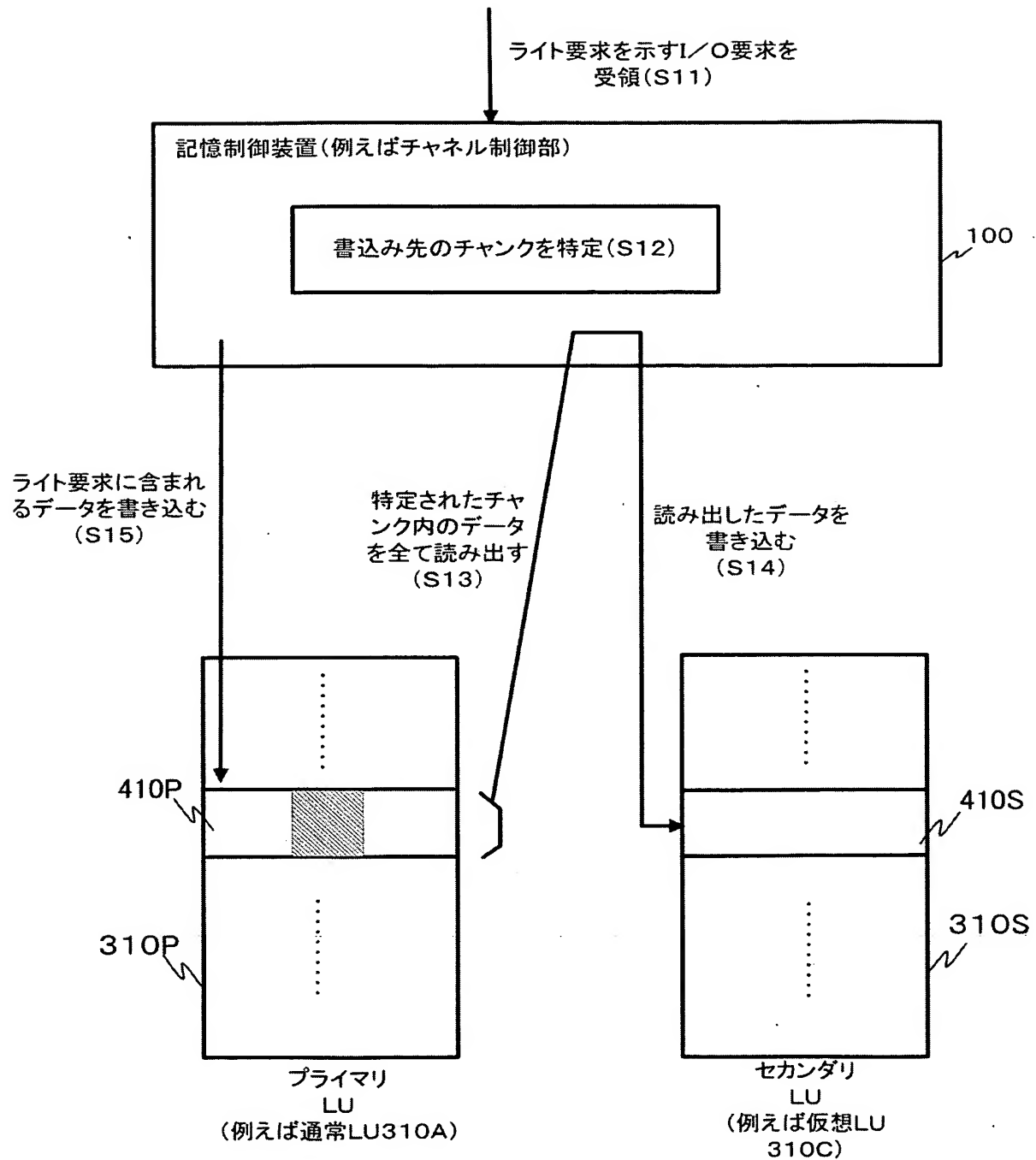
チャンネル制御部の動作



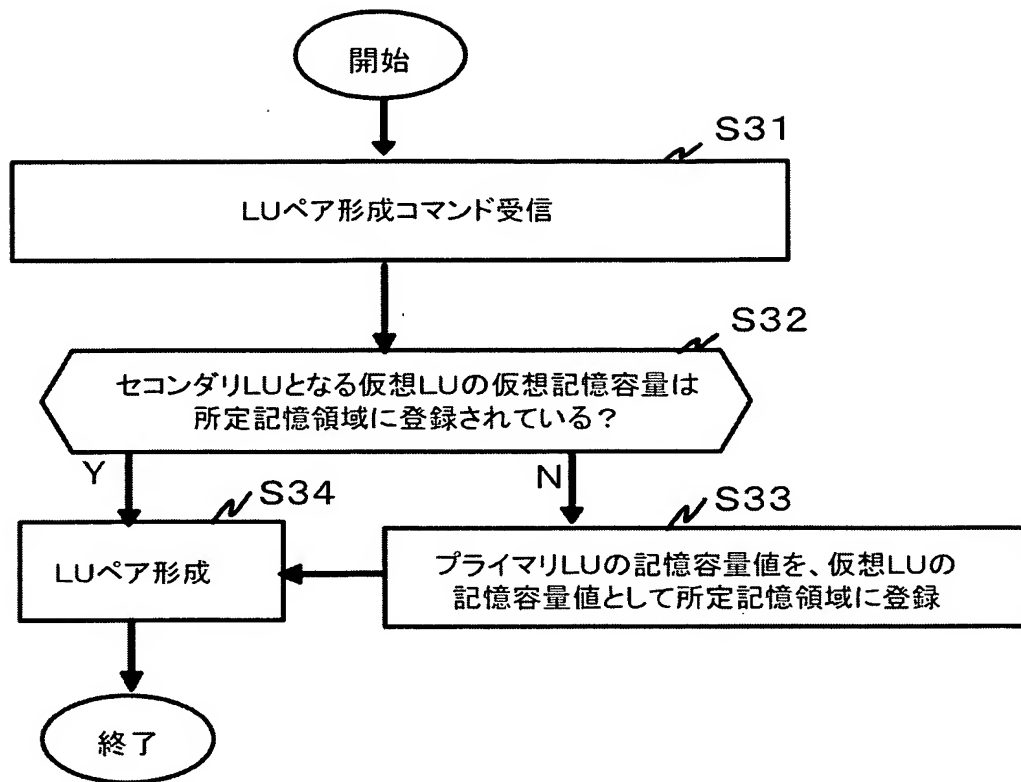
【図 12】



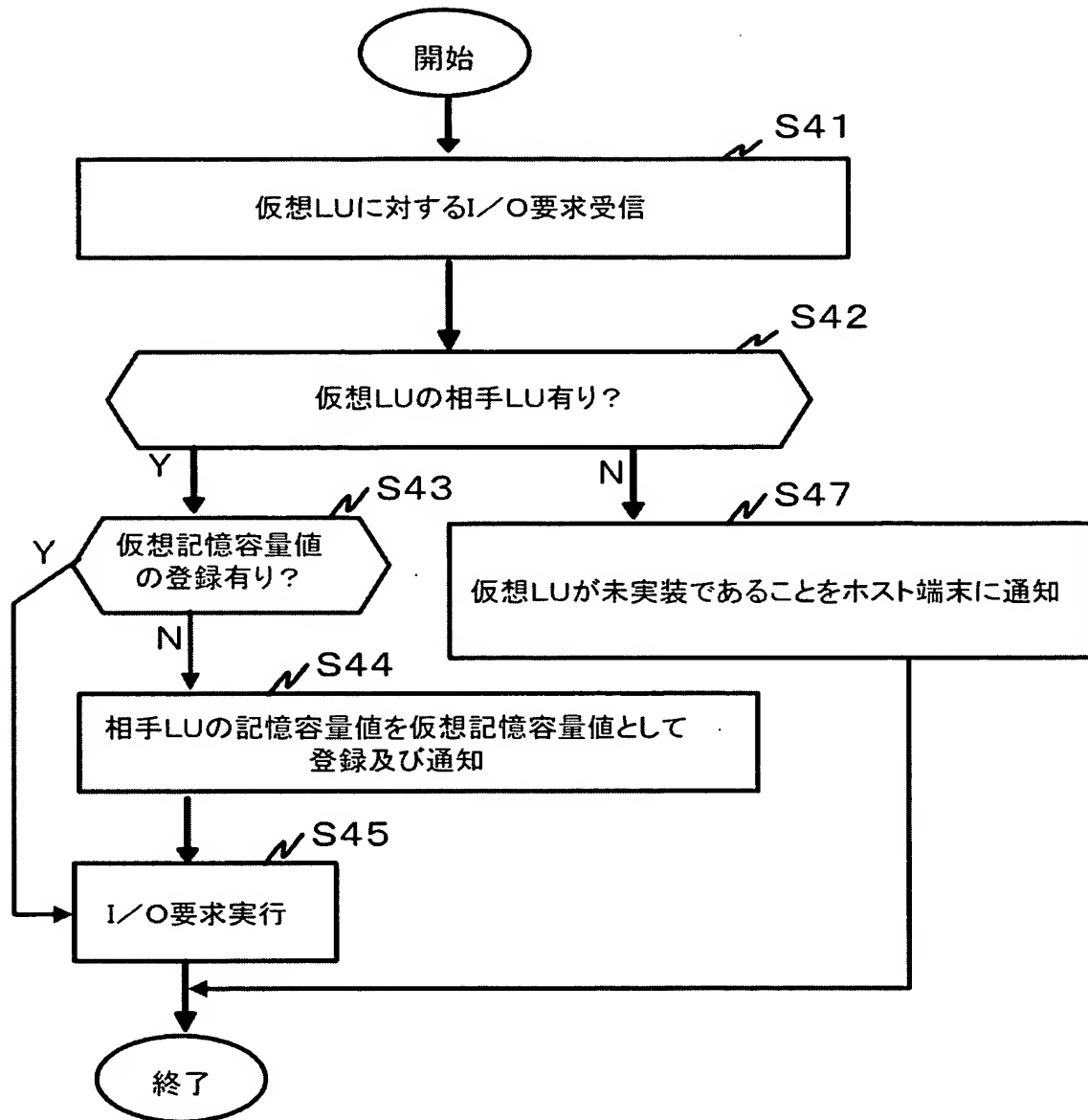
【図13】



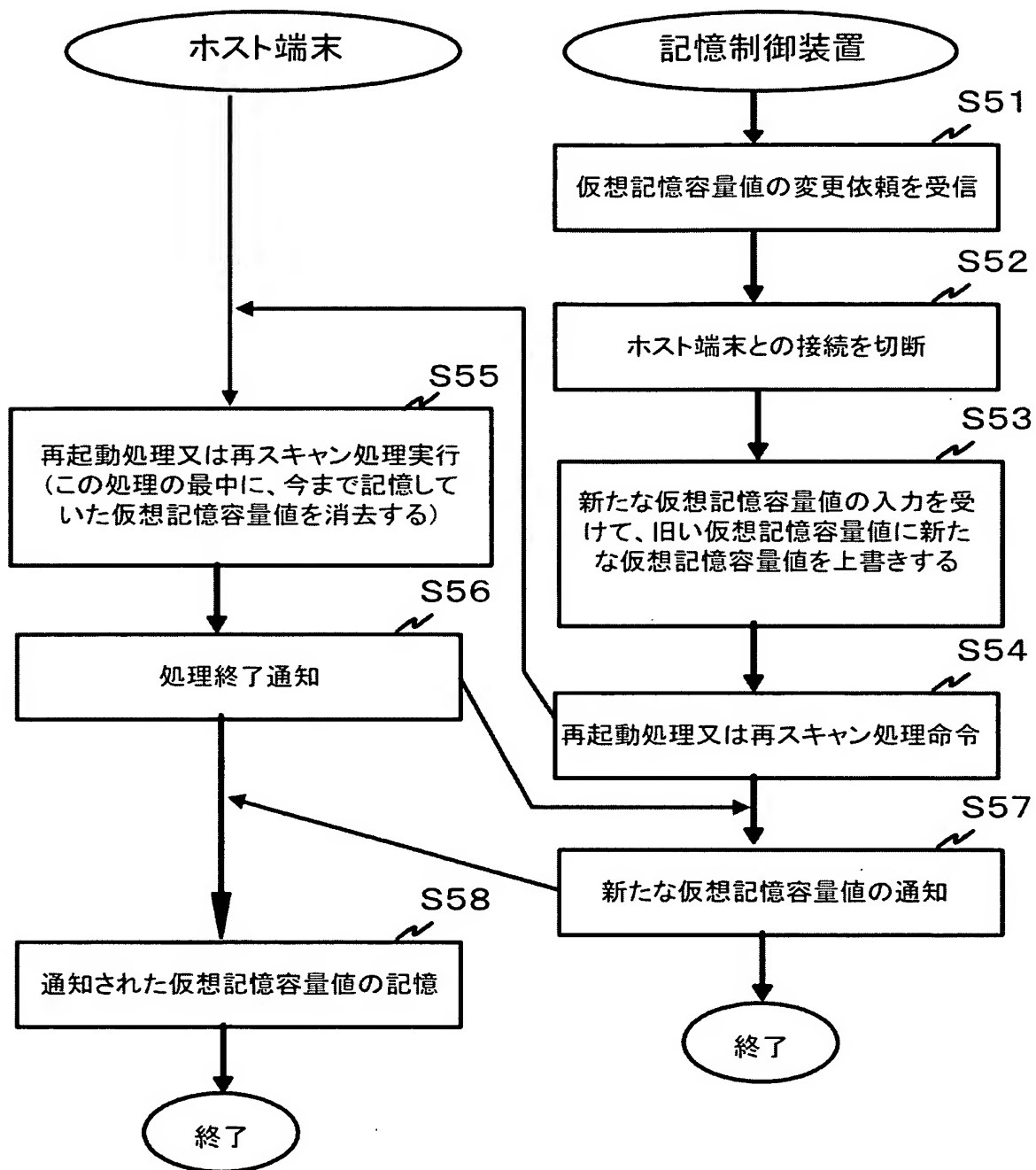
【図 14】



【図15】



【図 16】



【書類名】 要約書

【要約】

【課題】 ホスト端末に混乱を生じさせてしまう可能性を低減する記憶制御サブシステムが提供する。

【解決手段】 記憶制御サブシステム 1 0 2 に備えられる記憶制御装置 1 0 0 は、共有メモリ 1 0 2 に記憶された仮想記憶容量値をホスト端末に通知し、そのホスト端末において仮想記憶容量値が記憶された後、その仮想記憶容量値を有する仮想記憶ユニットがホスト端末に接続されている間は、上記通知した仮想記憶容量値が変更されないようにする。

【選択図】 図 2

認定・付加情報

特許出願の番号	特願 2004-031150
受付番号	50400201260
書類名	特許願
担当官	第七担当上席 0096
作成日	平成16年 2月 9日

<認定情報・付加情報>

【提出日】 平成16年 2月 6日

特願 2 0 0 4 - 0 3 1 1 5 0

出 願 人 履 歴 情 報

識別番号 [0 0 0 0 0 5 1 0 8]

1. 変更年月日	1 9 9 0 年 8 月 3 1 日
[変更理由]	新規登録
住 所	東京都千代田区神田駿河台 4 丁目 6 番地
氏 名	株式会社日立製作所